Separation Principles in Independent Process Analysis

Zoltán Szabó

Ph.D. Thesis

Eötvös Loránd University
Faculty of Informatics
Department of Information Systems

Supervisor: Prof. habil. András Lőrincz

Graduate School of Computer Science
János Demetrovics D.Sc.

Information Systems Program
András Benczúr D.Sc.

Budapest, 2008.

Contents

Acknowledgements

First of all, I would like to express my thanks to my supervisor András Lőrincz for the continuous support and trust. I'm deeply indebted to the members of his research group for the friendly atmosphere and for their readiness to help. Namely, to Zoltán Bárdosi, Ákos Bontovics, Viktor Gyenes, Márton Hajnal, György Hévízi, Melinda Kiszlinger, Katalin Lázár, Zsolt Palotai, Barnabás Póczos, Botond Szatmáry, Balázs Szendrő, Gábor Szirtes, István Szita and Bálint Takács. I'm especially grateful to Barnabás Póczos for our coloquies. I would also like to thank my teacher all the knowledge I gained over the years. I would like to say thanks to the US Air Force (EOARD), to the Neumann János Society, to the Bliss Foundation and to the Department of Information Systems for the financial support of my work.

I owe my parents thank for helping me during my studies and for all what a peaceful family life can provide. Thanks to my sister and to her little naughty children for always brightening me up with their joy and playfullness.

Chapter 1

Motivation

A natural requirement in signal processing, data mining, time series analysis and in the application of dynamical systems is to *decompose* the problem to be solved to independent subproblems. Such a well-known principle of system theory is the so-called *separation principle*:

- In a *stochastic control* problem we say that the separation principle holds if the optimal control *separates* into two stages based on optimal filtering of the unobservable state and optimal control of the filtered state. The simplest case when the separation principle holds, is the case of discrete time linear systems governed by Gaussian noise with quadratic objective, the LQG (Linear Quadratic Gaussian) problem [100, 96, 81][1]. A proof for the separation property in continuous time systems was given by [179], who extended the validity also for problems with non-quadratic cost. Since then, numerous separation principles concerning the optimal stochastic control have been proved, e.g., similar principles exist for linearly constrained LQG problems [120, 121], distributed parameter systems [31, 32], 'LQ-fractional Brownian motion' [111] and 'LQ-Hidden Markov Model' [151] problems, Wiener and polynomial systems [79], integral Volterra systems [28], the H_2-control problem of Markov jump linear systems [56, 57] and quantum control problems [37, 38]. A recent result is that the separation principle also holds in case of unknown parameters for LQG problems [27].

- In *stabilizing control* problems, where the task is to the construct a stabilizer

[1] The idea of separation has been early introduced to economics in [158, 167].

controller without the availability of the system state, similar separation theorems hold. The basic linear setting [123] can be extended, e.g., to nonlinear jump systems [132], distributed dissipative bilinear systems [36], non-uniformly completely observable systems [126] or switching linear systems [35]. The problem can also be broken into two separate parts in the closely related *robust pole placement* task [162, 80].

One of the most popular field in signal processing, the so-called Independent Component Analysis (ICA) [98, 55] may seem to have little in common with the *control problems* listed above. ICA can be considered as the 'nickname' of a *cocktail party* problem: there are D pieces of one-dimensional (independent) sound sources and D microphones and the task is to separate the original sources from the observed mixed signals. Independent Subspace Analysis (ISA), an extension of ICA, allows multidimensional components, too. One of the most exciting and fundamental hypotheses of the ICA research is due to Jean-François Cardoso who conjectured—by numerical experiments—that the solution of the ISA problem can be *decomposed* [41]: (i) one may set aside that there are subspaces in the background and invoke a classical ICA algorithm, then (ii) cluster the estimated ICA elements into statistically dependent groups. This principle can be considered as the *analogy* of the separation principles of system theory–hence we will refer to this *conjecture* as the *ISA separation principle*.

Provided that the ISA separation principle holds, non-combinatorial identification and realization of dynamical systems —that is the estimation of the parameters and the hidden variables—with independent multidimensional non-Gaussian variables may be attainable.[2] Up to now, neither this conjecture, nor its consequences has been justified completely. The present thesis is a step towards this direction.

Particularly, we prove reduction techniques, *separation principles*, which extend the i.i.d. (independent identically distributed) ISA separation principle to dynamical systems. The derived principles:[3]

- cover and generalize the classical assumptions of the ICA literature: multi-

[2]The combinatorial difficulty stems from the fact that, in the general case, the dimension of the independent subsystems (components) is not known. This difficulty can often be alleviated efficiently by non-combinatorial approximations (for details, see Section 5.3).

[3]We note, that the results of the present thesis can be extended to *controlled* dynamical systems driven by independent, multidimensional, non-Gaussian variables. However, the aim of this thesis is to present the theory of systems *without control*.

dimensional components; post nonlinear-, autoregressive-, convolutive mixing; complex-valued variables (Chapter 4).

- can be used to construct efficient and large-scale ISA methods (Chapter 5: Section 5.1-5.2).
- make it possible to estimate the dimension of the hidden components in ISA-reducible models by non-combinatorial approximations (Chapter 5: Section 5.3).

Before doing so, in Chapter 2 we give a brief introduction to independent component analysis. Chapter 3 formulates the problem domain and its ambiguities. Chapter 6 provides numerical experiments illustrating the efficiency of the algorithms built upon on the derived separation principles. Chapter 7 contains discussion and open questions. Proofs (roman numbers) and pseudocodes are put to Appendix A and B, respectively. Abbreviations used throughout the paper are listed in Appendix C. A short summary of the thesis is given in English and Hungarian after the Appendix on page 90 and 92, respectively.

Chapter 2

Overview of the ICA Related Literature

At the beginning I attempt to give a brief overview of the literature related to the search for independent components. Section 2.1 starts with the simplest, so-called Independent Component Analysis (ICA) problem. The overview goes on by relaxing the assumptions of ICA, such relaxations are 'one-dimensional–multidimensional components', 'i.i.d–autoregressive time evolution', 'instantaneous–convolutive mixing', 'linear–post-nonlinear mixing', 'real–complex variables'. A common feature of these relaxations is that there exist separability results for them, their ambiguities can be characterized.[1] These directions will be unified in Chapter 3 with separation principles presented in Chapter 4. The overview ends with enumerating other extensions of ICA not treated in this thesis (Section 2.2).

In what follows I focus on problems with hidden, independent, multidimensional, non-Gaussian variables. For better understanding, some formulae are given, but the exact definitions will be detailed in Section 3.1.2.

[1]Expressions *separability* and *separation* shouldn't be confused: separability refers to the ambiguities of the problem, separation is used as a synonym of a reduction step.

2.1 'Classical' Extensions of ICA

The simplest model dealing with independent components is Independent Component Analysis (ICA) [98, 55]. One can think of the ICA problem as a *cocktail-party problem*, where there are one-dimensional sound sources and microphones, and the task is to recover the original sources from the observed mixed signals. Formally, only the instantaneous linear mixture of hidden independent sources

$$\mathbf{x} = \mathbf{A}\mathbf{e} \tag{2.1}$$

is available for observation, where $\mathbf{A}$ is the mixing matrix, $\mathbf{e}$ is the hidden source to be estimated with independent $e_i \in \mathbb{R}$ coordinates. In spite of its simplicity, this model has been successfully applied to feature extraction [29], (ii) denoising [87], (iii) processing of financial [110] and neurobiological data, e.g., fMRI (functional Magnetic Resonance Imaging), EEG (ElectroEncephaloGraphy), and MEG (MagnetoEncephaloGraphy) [127, 176], and (iv) face recognition [26]. For a recent review about ICA see [54, 90, 53]. Nonetheless, applications in which the ICA conditions (*unechoic cocktail party, with independently talking participants and linearly recording microphones*—more precisely: linear, instantaneous mixing; one-dimensional, i.i.d hidden components) are not met may be highly relevant in practice:

Multidimensional, Autoregressive Components: For instance, consider the generalization of the cocktail-party problem, where *independent groups of people* are talking about independent topics or more than one *group of musicians* is playing at the party. The task requires an extension of ICA, which can be found under different names: Independent Subspace Analysis (ISA) [88], Independent Feature Subspace Analysis (IFSA) [108], Multidimensional Independent Component Analysis (MICA) [41], Subspace ICA [155] and group ICA [170]. We will use the first of these abbreviations throughout this paper. Formally, our goal is to estimate the independent hidden $\mathbf{e}^m \in \mathbb{R}^{d_m}$ components ($\mathbf{e} = [\mathbf{e}^1; \ldots; \mathbf{e}^M]$) from their instantaneous linear mixture

$$\mathbf{x} = \mathbf{A}\mathbf{e}. \tag{2.2}$$

The large number of different ISA algorithms [41, 15, 88, 91, 177, 24, 161, 146, 145, 170, 171, 172, 3, 137, 9, 117, 155, 45, 135, 136, 134, 157, 156, 50, 47] shows the importance of this field. The pioneering work of [41] (i) is based on geometric considerations, and (ii) poses the possibility to solve the ISA problem as the grouping of ICA elements. We will return later to this fundamental idea, the *ISA separation principle* in Chapter 4 and to its simbling emerged at the Joint Block Diagonalization (JBD) community. The hidden sources are modelled by forests with a cost functions made of mutual information of the estimated coordinates in the work of [24]. The ISA problem is formulated as the optimization of multidimensional differential entropies in [146, 145], where the estimation of the entropy is carried out by the use of k-nearest neighbors and geodesic spanning tree methods, respectively. Mutual information and entropy based ISA cost functions (joint/pairwise, multidimensional/one-dimensional) are derived in [9], and the Kernel Canonical Correlation Analysis, Kernel Generalized Variance (KGV) [23], and the Kernel Covariance (KC) [78] techniques are generalized for the estimation of mutual information of multidimensional varibles. The Joint $\mathcal{F}$-Decorrelation (JFD) ISA method of [3] aims to decorrelate (block diagonalize) over a function set. The ISA problem is formulated as JBD in [170, 171, 172], too. Namely, [171, 170] aims at the joint block diagonalization of the Hessian of (i) characteristic functions, (ii) logarithmic densities, respectively; the MSOBI (Multidimensional SOBI; Multidimensional Second-Order Blind Identification) technique executes temporal joint block-decorrelation, and a cumulant based solution is shown in [172]. The fastICA method [92] is generalized to ISA in [91]. The ISA task is transformed to Maximum Likelihood (ML) estimation for spherically symmetric hidden source components in [88]. Recent manifold optimalization techniques are adapted to this ISA cost function in [135, 136, 134, 157, 156, 50]. Other recent approaches search for independent subspaces via (i) hierarchical mutual information based clustering [161], (ii) cumulant based objective function [47], (iii) separability in the phase space applying differential geometric tools [117] and (iv) vector kurtosis based ISA cost [155]. A flexible component model framework is developed by [45], which optimizes the matching of covariance matrices, and can give totally blind, semi-blind and non-blind procedures depending on the chosen parameterization of the covariance matrices.

Successful applications of ISA involve: (i) the processing of EEG-fMRI data [15], (ii) gene analysis [108, 107, 109], (iii) face view recognition [118, 119], (iv) ECG (ElectroCardioGraphy) analysis [161, 170].

Temporal independence of ISA is, however, a gross oversimplification of real sources including acoustic or biomedical data. One may try to overcome this problem, by assuming that the hidden processes are, e.g., AutoRegressive (AR) processes. Then we arrive at the AR Independent Process Analysis (AR-IPA) task [141, 86, 49, 148, 6, 84, 48, 19, 51, 147, 1, 95, 8, 5]:

$$\mathbf{s}(t) = \sum_{i=1}^{L_s} \mathbf{F}_i \mathbf{s}(t-i) + \mathbf{e}(t), \tag{2.3}$$

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \tag{2.4}$$

Here, (i) the hidden process $\mathbf{s}$ is driven by a non-Gaussian variable $\mathbf{e}$ satisfying the assumptions of the ISA problem and (ii) the linear mixture $\mathbf{x}$ is observed.

One-dimensional hidden components: ML based algorithms are derived for the context-sensitive ICA problem [141], and by assuming generalized exponential innovation [49] The AR-IPA task is reduced to applying ICA to the innovation process produced by linear prediction (AR fit) [86]—sometimes temporal differentiating is enough instead of AR estimation [51]. An expectation maximization technique is derived under the assumption of mixture of Gaussian innovation [84], and the innovation is estimated in ML framework in [48].

Multidimensional hidden components: The innovation trick is extended to the case of multidimensional hidden components in [148]. The Cross-Entropy (CE) method [153] can be tailored to perform ISA on the innovation [6]. Using the ISA separaration theorem [9], the hidden one-dimensional processes are clustered by the predictive matrix of the hidden source, and the information matrix of the estimated innovation in [147, 1, 8] and [5], respectively. The Independent Dynamics Subspace Analysis method aims at minimizing the error (in L_2 sense) of the difference process of the estimated hidden source. Multidimensional, hidden sources

are estimated in [19], too. However, the model behind their solution is different. Although the linear mixture is assumed to be constrained, in contrast to AR, more general (stationary and ergodic) hidden processes can be assumed.

Temporal Mixing (Convolution): Another extension of the original ICA task is the Blind Source Deconvolution (BSD) problem. Such a problem emerges, for example, at a cocktail-party being held in an *echoic room*, and can be modelled by a convolutive mixture

$$\mathbf{x}(t) = \sum_{j=0}^{L_e} \mathbf{H}_j \mathbf{e}(t-j), \tag{2.5}$$

where the $\mathbf{e}$ hidden variable has $e_m \in \mathbb{R}$ independent coordinates.

Several BSD algorithms have been developed over the last decades, for a review see [142]. BSD shows potentials in the following areas: (i) remote sensing applications; passive radar/sonar processing [125, 83], (ii) image-deblurring, image restoration [178], (iii) speech enhancement using microphone arrays, acoustics [65, 130, 152, 22], (iv) multi-antenna wireless communications, sensor networks [16, 61], (v) biomedical signal—EEG, ECG, MEG, fMRI—analysis [97, 74, 66], (vi) optics [113], (vii) seismic exploration [102].

Nonlinear Mixing: The strong assumption on the linearity can be relaxed by assuming component-wise distortion resulting in a Post NonLinear extension (PNL-ICA) of the ICA [165]. In PNL-ICA, the observation is

$$\mathbf{x} = \mathbf{f}(\mathbf{A}\mathbf{e}), \tag{2.6}$$

where function $\mathbf{f}$ acts component-wise. This direction has recently gained much attention, for a review see [99].

Complex variables: The ICA research has been concerned with complex variables from its early years [46, 55]. Since then, a number of $\mathbb{C}$-ICA methods has appeared [21, 25, 72, 33, 55, 46, 42, 131, 44, 73, 30, 181, 159, 39, 180, 69, 63, 64, 134, 124, 59, 43, 138, 139, 116, 62]. Possible reasons for that:

- The application areas of Complex-Valued Neural Networks (CVNN) have widened recently. For an excellent review about CVNNs consult [85].
- Particularly, there is a natural tendency to apply complex-valued computations for the analysis of biomedical signals. Complex ICA ($\mathbb{C}$-ICA) approximation has been applied for the analysis of EEG [21] and fMRI [40, 20] data.

The main lines of the $\mathbb{C}$-ICA techniques are the followings: ML principle and complex recurrent neural network are used in [21, 43], and [25], respectively. Nongaussianity is maximized in [138]. The Adaptive Principal component EXtractor (APEX) algorithm is based on Hebbian learning [72]. Complex FastICA algorithm can be found in [33]. More solutions are based on cumulants: e.g., [55, 124], the Joint Approximate Diagonalization of Eigen-matrices (JADE) algorithm [46, 42], its higher order variants [131], and the Equivariant Adaptive Separation via Independence (EASI) algorithm family [44]. 'Rigid-body' learning theory is used in [73], [134] performs optimization on the flag manifold. The SOBI algorithm [30] searches for joint diagonalizer matrix, its refined version, the weights-adjusted SOBI method [181] approximates by means of weighted nonlinear least squares. There are complex variants of the infomax (information maximization) technique, such as the a split-complex [159] and the fully-complex infomax [39] procedures. The estimation of a $\mathbb{C}$-ICA coordinate is carried out in a deflation framework by the minimization of the support of the coordinate [59]. [139] presents a demixing technique for Quadrature Amplitude Modulated (QAM) sources using gaussian mixture model. The spacing idea of the Robust, Accurate, Direct ICA aLgorithm (RADICAL) [115] is adapted to the absolute value of circular sources whose density depends only on the absolute value of the argument [116]. Minimax mutual information [180] and Strong-Uncorrelating Transforms (SUTs) [69, 63, 64, 62] represent other promising directions.

It is important to see to what extent we can expect to regain the true sources. A common property shared by the above problems (ICA, ISA, AR-IPA, BSD, PNL-ICA) is that *their separability has been proven.* In other words, these problems are well-defined, their ambiguities (indeterminacies) can be characterized. Furthermore,

as we will see later (Chapter 4) it is sufficient to consider the ambiguities of the ICA, PNL-ICA and ISA problems as all other cases can be derived from these basic types.

- In ICA, hidden sources can be recovered up to a scalar multiplier and permutation [55, 68].
- In addition to these ambiguities there is an additive scalar term in the PNL-ICA problem [14].
- In the ISA problem, the components of equal dimension can be recovered up to the permutation (of equal dimension) and invertible transformation within the subspaces [169, 172].

2.2 Other Extensions of ICA

The following problems represent other important directions in the field of source separation concerning hidden multidimensional non-Gaussian components:

Non-Gaussian Component Analysis, Colored Subspace Analysis: The non-Gaussian Component Analysis (NGCA) [34, 173, 105, 106, 104] and the Colored Subspace Analysis (CSA) [174] are such relatives of the ISA problem, where ambiguities can still be determined.

- In the NGCA problem, the observation is instantanenous linear mixture of two independent i.i.d. components (one Gaussian and one non-Gaussian)

$$\mathbf{x} = \mathbf{A}[\mathbf{e}_{\text{non-Gaussian}}; \mathbf{e}_{\text{Gaussian}}] \tag{2.7}$$

  and the goal is to estimate the non-Gaussian signal subspace ($\mathbf{e}_{\text{non-Gaussian}}$). In contrast to ISA however, no assumption of independence within the non-Gaussian hidden signal subspace is made.

- In CSA, the observation is the instantaneous linear mixture of two hidden 'independent' (their auto-crosscorrelation vanishes) components (an i.i.d, and a Wide-Sense Stationary (WSS)), processes

$$\mathbf{x} = \mathbf{A}[\mathbf{s}_{\text{WSS}}; \mathbf{s}_{\text{iid}}], \tag{2.8}$$

and the task is to project to the 'colored subspace', that is to estimate $\mathbf{s}_{\mathrm{WSS}}$.

Topographic/Hierarchical Organization of the Components: Topographic ICA (TICA) is ICA dressed up with topographic organization: the energy of the coordinates is allowed to be correlated in their neighbourhood. The TICA task can be formulated as an ML estimation, and reduces to ISA upon a special choice of the neighbourhood structure. A 2-layer neural network is applied for feature extraction by [112]: the first layer performes ICA (precisely, is initialized by ICA) and the elements of the first layer are grouped by the second layer–according to the numerical experiments. Similar hierarchical structure is presented in the [101] and in the Product of Experts (PoT) model [140].

Variance-Dependent Component Analysis: The dependence of the hidden coordinates (one-dimensional sources) are modelled in Variance-Dependent Component Analysis (VDCA) [89, 103, 93] by the dependence of their time-evolving variances.

Chapter 3

The Independent Process Analysis Problem

In this chapter, I present a unification of the ICA extensions of Section 2.1 (ISA, AR-IPA, BSD, PNL-ICA and complex models) by recoursing to Integrated ARMA (ARIMA) processes, which are commonly used for economic phenomena [129]. Definitions are given in Section 3.1, ambiguity questions are addressed in Section 3.2.

3.1 Definitions

To put our ARIMA Independent Process Analysis (ARIMA-IPA) model into mathematical form (Section 3.1.2), we need a few notations (Section 3.1.1).

3.1.1 Polynomial Matrices, Complex Random Variables

Numbers (v), vectors ($\mathbf{v}$), and matrices ($\mathbf{V}$) are denoted by different letter types. $\mathbb{N} = \{0,1,\dots\}$ stands for natural numbers. $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ may stand for either real or for complex numbers. Let $\mathbb{K}^{D_1 \times D_2}$ be the set of $D_1 \times D_2$ matrices over $\mathbb{K}$. Let z stand for the time-shift operation, that is $(z\mathbf{v})(t) := \mathbf{v}(t-1)$. Polynomials of $D_1 \times D_2$ matrices are denoted by $\mathbb{K}[z]^{D_1 \times D_2} := \{\mathbf{V}[z] = \sum_{n=0}^{N} \mathbf{V}_n z^n, \mathbf{V}_n \in \mathbb{K}^{D_1 \times D_2}\}$. A $\mathbf{V}[z] \in \mathbb{K}[z]^{D_1 \times D_2}$ polynomial matrix maps a series of vectors $\{\mathbf{v}(t) \in \mathbb{K}^{D_2}\}$ to $(\mathbf{V}[z]\mathbf{v})(t) = \sum_{n=0}^{N} \mathbf{V}_n \mathbf{v}(t-n)$. Let $\nabla^r[z] := (\mathbf{I} - \mathbf{I}z)^r$ denote the operator of the r^{th} order difference ($r \in \mathbb{N}$), where $\mathbf{I}$ is the identity matrix. $Gl_{\mathbb{K}}(D)$ is the general lin-

ear group over $\mathbb{K}$: the set of invertible matrices from $\mathbb{K}^{D\times D}$ with the standard matrix product.

We introduce the basic concepts for using complex random variables. An excellent review on this topic can be found in [68]. $\mathbf{V}^T$ is the transposed of matrix $\mathbf{V} \in \mathbb{C}^{L\times L}$. Element-wise complex conjugation is denoted by bar. The transposed complex conjugate of matrix $\mathbf{V}$ is the adjoint matrix $\mathbf{V}^* = \bar{\mathbf{V}}^T$. Matrix $\mathbf{V} \in \mathbb{C}^{L\times L}$ ($\mathbb{R}^{L\times L}$) is called *unitary* (*orthogonal*) if $\mathbf{V}\mathbf{V}^* = \mathbf{I}$ ($\mathbf{V}\mathbf{V}^T = \mathbf{I}$). The sets of $L \times L$ dimensional unitary and orthogonal matrices are denoted by $\mathcal{U}^L$ and $\mathcal{O}^L$, respectively. A *complex-valued random variable* $\mathbf{v} \in \mathbb{C}^L$ (shortly complex random variable) is defined as a random variable of the form $\mathbf{v} = \mathbf{v}_R + i\mathbf{v}_I$, where the real and imaginary parts of $\mathbf{v}$, i.e., $\mathbf{v}_R$ and $\mathbf{v}_I \in \mathbb{R}^L$ are real random variables, $i = \sqrt{-1}$. Expectation value of complex random variables is $E[\mathbf{v}] = E[\mathbf{v}_R] + iE[\mathbf{v}_I]$, and the variable can be characterized in second order by its *covariance matrix* $cov[\mathbf{v}] = E[(\mathbf{v} - E[\mathbf{v}])(\mathbf{v} - E[\mathbf{v}])^*]$ and by its *pseudo-covariance matrix* $pcov[\mathbf{v}] = E[(\mathbf{v} - E[\mathbf{v}])(\mathbf{v} - E[\mathbf{v}])^T]$. Complex random variable $\mathbf{v}$ is called *full*, if $cov[v]$ is positive definite. Throughout this paper all complex variables are assumed to be full (that is, they are not concentrated in any lower dimensional complex subspace).

3.1.2 The ARIMA-IPA Model

The definition of the $\mathbb{K}$-ARIMA-IPA task as is follows. We assume M independent hidden random variables (*components*). There are $ARIMA(L_s, L_e, r)$ processes, where $L_e, L_s, r \in \mathbb{N}$, driven by these components, but only their linear mixture is available for observation. Formally,

$$\mathbf{F}[z]\nabla^r[z]\mathbf{s} = \mathbf{H}[z]\mathbf{e}, \tag{3.1}$$

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{3.2}$$

where $\mathbf{e}(t) = \left[\mathbf{e}^1(t); \ldots; \mathbf{e}^M(t)\right] \in \mathbb{K}^{D_e}$ ($D_e = \sum_{m=1}^{M} d_m$) is a vector concatenated of the independent components $\mathbf{e}^m(t) \in \mathbb{R}^{d_m}$.[1] The dimensions of observation $\mathbf{x}$ and hidden source $\mathbf{s}$ are D_x and D_s, respectively. $\mathbf{A} \in \mathbb{K}^{D_x\times D_s}$ is the so-called *mixing matrix*. $\mathbf{F}[z] := \mathbf{I} - \sum_{i=1}^{L_s} \mathbf{F}_i z^i \in \mathbb{K}[z]^{D_s\times D_s}$ and $\mathbf{H}[z] := \sum_{j=0}^{L_e} \mathbf{H}_j z^j \in \mathbb{K}[z]^{D_s\times D_e}$ are

[1] By d_m-dimensional $\mathbf{e}^m$ components, we mean that $\mathbf{e}^m$s cannot be decomposed into smaller dimensional independent parts. This property is called *irreducibility* in [172].

polynomial matrices that represent the AR and MA (Moving Average; also called convolution) parts, respectively. The goal of the $\mathbb{K}$-ARIMA-IPA task is (to estimate a *demixing system*) to recover the original source $\mathbf{e}(t)$ from observations $\mathbf{x}(t)$.

Our $\mathbb{K}$-ARIMA-IPA assumptions are listed below:

1. Dimensions: $D_x \geq D_s \geq D_e$.
2. Components:
   (a) for a given m, $\mathbf{e}^m(t)$ is i.i.d. in time t,
   (b) $I(\mathbf{e}^1, \ldots, \mathbf{e}^M) = 0$, where I stands for the mutual information of the arguments. This property will be referred to as $\mathbf{d}$-*independence* [$\mathbf{d} = (d_1, \ldots, d_M)$].
   (c) There is at most one Gaussian variable among $\mathbf{e}^m$. This assumption will be referred to as the 'non-Gaussian' assumption.[2]
3. $\mathbf{A} \in \mathbb{K}^{D_x \times D_s}$ has full column rank.
4. Polynomial matrix $\mathbf{F}[z]$ is *stable*, that is $\det(\mathbf{F}[z]) \neq 0$, for all $z \in \mathbb{C}, |z| \leq 1$.
5. Polynomial matrix $\mathbf{H}[z]$ has *left inverse*. In other words, there exists a polynomial matrix $\mathbf{Q}[z] \in \mathbb{R}[z]^{D_e \times D_s}$ such that $\mathbf{Q}[z]\mathbf{H}[z] = \mathbf{I}$.

Specially, for $r = 0$, the $\mathbb{K}$-ARMA-IPA special task appears:

$$\mathbf{F}[z]\mathbf{s} = \mathbf{H}[z]\mathbf{e}, \tag{3.3}$$

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \tag{3.4}$$

Let us note that the stability of $\mathbf{F}[z]$ implies the stationarity of ARMA process $\mathbf{s}$. The core difference for $0 \neq r$ is that Eq. (3.1) includes r^{th} order differentiating, which makes the process non-stationary. Special $\mathbb{K}$-ARMA-IPA tasks are the followings:

[2] In the complex case, this non-Gaussian constraint can be relaxed: the mixture of certain Gaussian variables can also be demixed, see Section 4.4.

1. The case of $L_e = 0$ corresponds to the $\mathbb{K}$-AR-IPA task[3]:

$$\mathbf{s}(t) = \sum_{i=1}^{L_s} \mathbf{F}_i \mathbf{s}(t-i) + \mathbf{e}(t), \tag{3.5}$$

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \tag{3.6}$$

   If $L_s = 0$ also holds, then the $\mathbb{K}$-ISA (or $\mathbb{K}$-IID-IPA) task emerges:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{e}(t) \tag{3.7}$$

   In addition, if d_m is chosen to be one—that is the $\mathbf{e}^m$ components are one-dimensional—then the task reduces to the simpler $\mathbb{K}$-ICA problem.

2. for $L_s = 0$, one talks about the $\mathbb{K}$-MA-IPA (or $\mathbb{K}$-Blind Subspace Deconvolution, $\mathbb{K}$-BSSD) task[4]:

$$\mathbf{x}(t) = \sum_{j=0}^{L_e} \mathbf{H}_j \mathbf{e}(t-j) \tag{3.8}$$

   In other words, the causal FIR (Finite Impulse Response) filtered mixture[5] of the hidden components is available for observation. Furthermore, if $\forall d_m = 1$ then we arrive at the original $\mathbb{K}$-BSD problem.

The relations among the different tasks are summarized in Fig. 3.1. For $D_x > D_e$ the problem is called *undercomplete*, while the case of $D_x = D_e$ is regarded as *complete*. We are dealing with the complete problem, if it is not stated otherwise; to distinguish the undercomplete case, prefix 'u' is used, e.g., uARIMA-IPA.[6] For real variables ($\mathbb{K} = \mathbb{R}$), we address the problem of post nonlinear (PNL) models. In this case, observation equation (3.2) is replaced by

$$\mathbf{x} = \mathbf{f}(\mathbf{A}\mathbf{s}), \tag{3.9}$$

[3] In $\mathbb{K}$-AR-IPA, besides $\mathbf{e}$ quantities $\mathbf{s}, \mathbf{F}[z], \mathbf{A}, \mathbf{A}^{-1}$ are also estimated.

[4] In $\mathbb{K}$-MA-IPA, besides $\mathbf{e}$, $\mathbf{H}[z]$ can also be easily estimated by the help of $\mathbf{x}$, e.g., by using least squares.

[5] Causal: $j \geq 0$ in $\sum_j$. FIR: the number of terms in the sum is finite.

[6] One can show for $D_s > D_e$ that under mild conditions $\mathbf{H}[z]$-has an inverse with probability 1 [150]; e.g., when the matrix $[\mathbf{H}_0, \ldots, \mathbf{H}_{L_e}]$ is drawn from a continuous distribution (see assumption 5, on page 15).

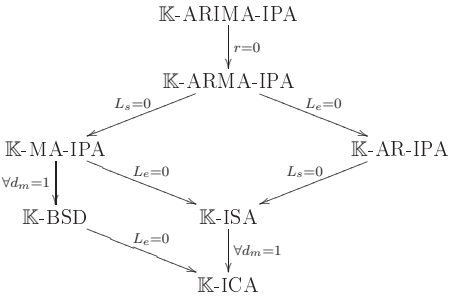


Figure 3.1: The general $\mathbb{K}$-ARIMA-IPA model and its special cases. Arrows point to special cases. For example, '$\mathbb{K}$-ISA $\xrightarrow{\forall d_m=1}\mathbb{K}$-ICA' means, that the $\mathbb{K}$-ICA model is a special case of the $\mathbb{K}$-ISA model, when all components are one-dimensional.

where function $\mathbf{f} = [f_1; \ldots; f_{D_x}] : \mathbb{R}^{D_x} \to \mathbb{R}^{D_x}$ is a component-wise transformation that is $\mathbf{f}(\mathbf{x}) = [f_1(x_1); \ldots; f_{D_x}(x_{D_x})]$ and $\mathbf{f}$ is invertible. Function $\mathbf{f}$ can be conceived as a distortion acting component-wise.

3.2 Ambiguities

This section is about the ambiguity (also called separability) issues of ISA models: what extent is it possible recover the original hidden components. The linear ISA case ($\mathbb{R}$-ISA, $\mathbb{C}$-ISA) is detailed in Section 3.2.1, the PNL-ISA case is addressed in Section 3.2.2. As we will see it later (Chapter 4), other cases in the $\mathbb{K}$-ARIMA-IPA problems can be transformed into one of these problems.

3.2.1 $\mathbb{K}$-ISA Separability

Identification of the (complete) $\mathbb{K}$-ISA model is ambiguous. However, the ambiguities of the model are simple: hidden components can be determined up to permutation of the subspaces (of equal dimension) and invertible transformation within the subspaces. The proof of the real case can be found in [169, 172], the non-Gaussian complex case can be transformed to the real case by the technique detailed in Section 4.4 (see

Proof I). Ambiguities within subspaces can be lessened. One may assume without any loss of generality that both the observed ($\mathbf{x}$) and the hidden ($\mathbf{e}$) signals are *white*: their (i) expected values are zero, (ii) covariance matrices are identity. Now,

- the mixing matrix $\mathbf{A}$ is orthogonal(real case)/unitary(complex case), because

$$\mathbf{I} = cov[\mathbf{x}] = E\left[\mathbf{x}\mathbf{x}^T\right] = \mathbf{A}E\left[\mathbf{e}\mathbf{e}^T\right]\mathbf{A}^T = \mathbf{A}\mathbf{I}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T, \tag{3.10}$$

$$\mathbf{I} = cov[\mathbf{x}] = E\left[\mathbf{x}\mathbf{x}^*\right] = \mathbf{A}E\left[\mathbf{e}\mathbf{e}^*\right]\mathbf{A}^* = \mathbf{A}\mathbf{I}\mathbf{A}^* = \mathbf{A}\mathbf{A}^*. \tag{3.11}$$

  holds in the real and complex case, respectively.

- thus, the $\mathbf{W} := \mathbf{A}^{-1}$ demixing matrix is orthogonal(real case)/unitary(complex case),

- and the $\mathbf{e}^m$ components are determined up to permutation (of equal dimension) *and* up to orthogonal(real case)/unitary(complex case) transformation within the subspaces.

The undercomplete ISA problem can be transcribed into complete ISA by applying Principal Component Analysis (PCA). In fact, the whitening and the undercomplete-complete reduction can be carried out in a single step by PCA (see Note II).

3.2.2 PNL-ISA Separability

To solve the PNL-ISA problem it is important to see to what extent we can expect to regain the true sources. For this generalization we prove a separability theorem (Theorem 1): the ambiguities of this problem are essentially the same as for the linear ISA task (Section 3.2.1). By applying this result we derive an algorithm using the mirror structure of the mixing system (Section 4.3).

Because of the PNL assumption the hidden sources can be estimated using the mirror structure of the mixing system, that is

$$\hat{\mathbf{e}} = \mathbf{W}\mathbf{g}(\mathbf{x}), \tag{3.12}$$

where $\mathbf{W} \in Gl_{\mathbb{R}}(D_e)$ and $\mathbf{g} : \mathbb{R}^{D_e} \to \mathbb{R}^{D_e}$ is a component-wise transformation. It has to be shown, however, that the $\mathbf{d}$-independence of the resulting $\hat{\mathbf{e}}$ unequivocally means

that the true $\mathbf{e}$ has been found. The following separability theorem shows that indeed this is the case. This statement (i) concerns the case of equal dimensions ($d = d_m, \forall m$) and (ii) can be considered as an extension of the results in [168] for the case $d \geq 1$. The proof of the theorem will be based on Lemmas 3.4 and 3.5 of [171]. These lemmas will only be cited. Let $C^2(V, \mathbb{R})$ and $C^\omega(V, \mathbb{R})$ denote the V (open) $\subseteq \mathbb{R}^n \to \mathbb{R}$ 2-times continuously differentiable and analytic functions, respectively. Let us introduce a concept: a matrix $\mathbf{U} \in \mathbb{R}^{dM \times dM}$ is called 'mixing', if there exist at least two invertible elements in any of its rows (considering $d \times d$ blocks as elements).[7] Formally, $\mathbf{U}$ is 'mixing', if decomposing matrix $\mathbf{U}$ into blocks of size $d \times d$ ($\mathbf{U} = [\mathbf{U}_{ij}]_{i,j=1,\ldots,M}, \mathbf{U}_{ij} \in \mathbb{R}^{d \times d}$), then for any index $i \in \{1, \ldots, M\}$ there exist a pair of indices $j \neq k \in \{1, \ldots, M\}$ for which matrices $\mathbf{U}_{ij}$ and $\mathbf{U}_{ik}$ are invertible. Our PNL-ISA separability result is the following:

Theorem 1 ([11]; PNL-ISA Ambiguities with Locally-Constant Nonzero C^2 Densities, $d = d_m(\forall m)$**)** *Let (i)* $\mathbf{A}, \mathbf{W} \in Gl_{\mathbb{R}}(D_e)$*, be 'mixing' matrices; (ii)* $\mathbf{e}$ *be as in* $\mathbb{R}$*-ISA with existing covariance matrix, and somewhere locally constant nonzero density function* $p_E \in C^2(\mathbb{R}^{D_e}, \mathbb{R})$*, (iii)* $\mathbf{h} : \mathbb{R}^{D_e} \to \mathbb{R}^{D_e}$ *is a component-wise bijection with coordinate functions in* $C^\omega(\mathbb{R}, \mathbb{R})$*. In this case, if* $\mathbf{y} := [\mathbf{y}_1; \ldots; \mathbf{y}_M] = \mathbf{Wh}(\mathbf{Ae})$ *is* $\mathbf{d}$*-independent (*$\mathbf{y}^m \in \mathbb{R}^d$*) with somewhere locally constant density function, then*

- $\mathbf{h}(\mathbf{x}) = \mathbf{Lx} + \mathbf{p}$*, where* $\mathbf{L} \in Gl_{\mathbb{R}}(D_e)$ *is a diagonal matrix and* $\mathbf{p} \in \mathbb{R}^{D_e}$*, and*
- *components* $\mathbf{y}^m$ *(*$m = 1, \ldots, M$*) recover the hidden* $\mathbf{e}^m$ *sources up to permutation and invertible transformation within the subspaces (and maybe up to a constant translation).*

[See Proof III on page 72.]

For example, uniform distributions belong to the family of somewhere locally constant and nonzero distributions, hence we have proven post nonlinear separability for uniformly distributed sources.

[7] It is easy to show an example, when it is impossible to estimate nonlinearities f_i and restore the source components $\mathbf{e}^m \in \mathbb{R}^d$, provided that the 'mixing' property is not fullfilled. Let $\mathbf{A} = blockdiag(\mathbf{A}^1, \ldots, \mathbf{A}^M)$, or more generally a block-permutation matrix made of $d \times d$ blocks (a synonym for the latter is *d-scaling* matrix [170, 171]).

Chapter 4

Separation Principles for Independent Process Analysis

In what follows, we present separation principles for the $\mathbb{K}$-ARIMA-IPA problem family of Section 3.1.2. By using these principles, we can decompose the solution of the original problem to simple(r) subproblems.

First, in Section 4.1 we review former, existing separation techniques, which address special cases of the problem class. Our contributions concerning real linear (post nonlinear) and complex linear models are detailed in Section 4.2 (Section 4.3) and Section 4.4, respectively.

4.1 Former Decomposition Principles in the ARIMA-IPA Problem Family

There are numerous separation techniques for special ARIMA-IPA problems in the literature:

- According to the ISA separation principle [41, 9] the solution of the ISA task requires an ICA preprocessing step followed by a suitable permutation of the ICA elements.
- By applying an AR identification procedure first, the AR-IPA task can be transformed to the ISA task [148] that extends the result of [86] to multidimensional

sources. In certain cases, simple temporal differentiating may be sufficient for the reduction step [51].

- The undercomplete BSD task can be reduced
  - to ISA by temporal concatenation of the observations [71], or
  - to ICA
    - by Minimal Polynomial Basis (MPB) [76],
    - by means of spatio-temporal decorrelation [52], or
    - by Linear Prediction Approximation (LPA) [94, 60, 77, 13].
- In case of one-dimensional components ($d_m = 1$), the solution of uARMA-IPA boils down to ICA:
  - after applying (MPB+) LPA [75],
  - after the solution of a linear equation system (LES) constructed from cumulant matrices [163].
- The PNL-ICA problem is solved using 'gaussianization' followed by ICA [183, 160].

Now, we extend these separation principles. Namely:

- We give sufficient conditions for the $\mathbb{R}$-ISA separation principle in Section 4.2.1. In Section 4.4.2, we justify its complex counterpart, the $\mathbb{C}$-ISA separation principle.
- We prove, that the $\mathbb{R}$-uMA-IPA task can also be reduced to $\mathbb{R}$-ISA by temporal concatenation, in the general $d_m \geq 1$ case (Section 4.2.2). Then, we show an alternative, linear prediction based principle, which is better suited for large scale problems.
- In Section 4.2.3, we reduce the $\mathbb{R}$-uARMA-IPA problem (in case of $d_m \geq 1$) to $\mathbb{R}$-ISA via linear prediction.
- The $\mathbb{R}$-uARIMA-IPA task can be solved by temporal differentiating followed by an $\mathbb{R}$-uARMA-IPA method (Section 4.2.4).

- The gaussianization based separation principle is extended to the PNL-ISA problem in Section 4.3.
- The complex linear models (up to $\mathbb{C}$-uARIMA-IPA) are transcribed to real problems in Section 4.4.1.[1]

The different separation principles (former+own) are summarized in Fig. 4.1.

4.2 Real Linear Models

In this section, separation principles for real linear models are elaborated.

4.2.1 The $\mathbb{R}$-ISA Separation Theorem - Sufficient Conditions

Without loss of generality, it can be assumed for an undercomplete/complete $\mathbb{R}$-ISA problem that it is complete (see Note II). According to the $\mathbb{R}$-ISA Separation Theorem, the $\mathbb{R}$-ISA problem can be solved by clustering the estimated $\mathbb{R}$-ICA elements into statistically dependent groups:

Theorem 2 ([9]; $\mathbb{R}$-ISA Separation Theorem) *Let* $\mathbf{y} = [y_1; \ldots; y_{D_e}] = \mathbf{W}\mathbf{x} \in \mathbb{R}^{D_e}$, *where* $\mathbf{W} \in \mathcal{O}^{D_e}$, $\mathbf{x} \in \mathbb{R}^{D_e}$ *is the whitened observation of the* $\mathbb{R}$*-ISA model, and* $D_e = \sum_{m=1}^{M} d_m$. *Let* $\mathcal{S}_{\mathbb{R}}^{d_m}$ *denote the surface of the* d_m*-dimensional unit sphere, that is* $\mathcal{S}_{\mathbb{R}}^{d_m} := \{\mathbf{w} \in \mathbb{R}^{d_m} : \sum_{i=1}^{d_m} w_i^2 = 1\}$. *Presume that the* $\mathbf{u} := \mathbf{e}^m \in \mathbb{R}^{d_m}$ *sources* $(m = 1, \ldots, M)$ *of the* $\mathbb{R}$*-ISA model satisfy condition*

$$H\left(\sum_{i=1}^{d_m} w_i u_i\right) \geq \sum_{i=1}^{d_m} w_i^2 H\left(u_i\right), \forall \mathbf{w} \in \mathcal{S}_{\mathbb{R}}^{d_m}, \tag{4.1}$$

and that the ICA cost function $J_{\mathbb{R}\text{-ICA}}(\mathbf{W}) = \sum_{i=1}^{D_e} H(y_i)$ *has minimum over the orthogonal matrices in* $\mathbf{W}_{\mathbb{R}\text{-ICA}}$. *Then it is sufficient to search for the solution to the* $\mathbb{R}$*-ISA task as a permutation of the solution of the* $\mathbb{R}$*-ICA task. Using the concept of*

[1] It is important to note that certain complex *methods* have already been rewritten to the real case (see, e.g., [43] and references therein). Here, we prove that the complex *problem* itself can be rewritten to a real one, thus *any* extant real procedure can be applied for the associated problem–provided that there is at least one hidden Gaussian component in the obtained real problem.

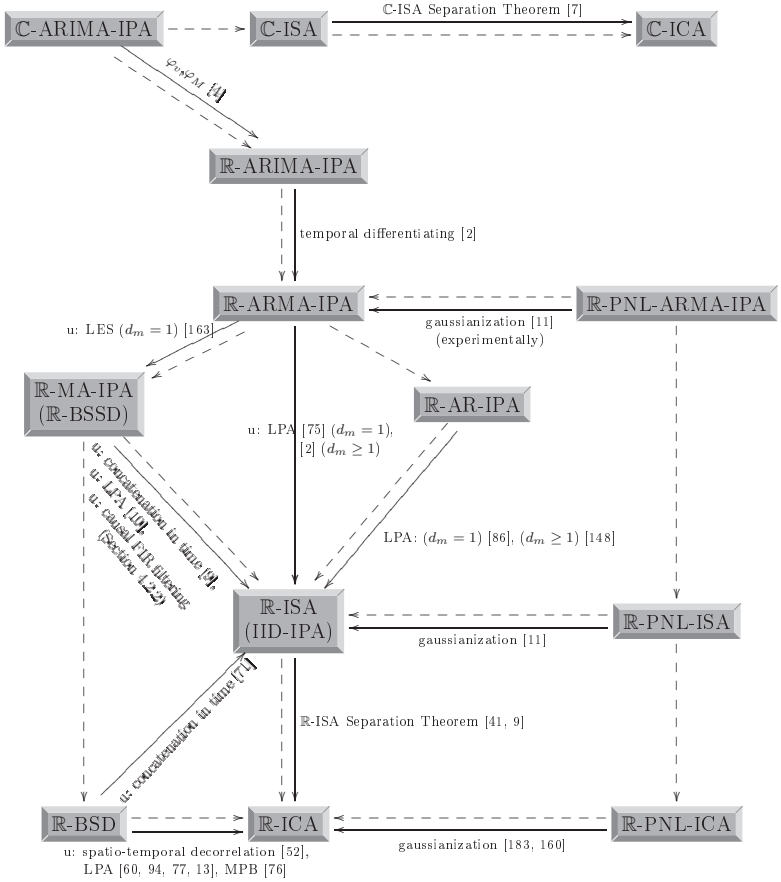


Figure 4.1: Illustration: the $\mathbb{K}$-ARIMA-IPA problem family and its reduction possibilities. Dashed arrows point to special cases. Solid arrows: represent separation principles. Prefix 'u': undercomplete case. $\mathbb{R}$ ($\mathbb{C}$): real (complex) case.

demixing matrices, it is sufficient to explore forms

$$\mathbf{W}_{\mathbb{R}\text{-ISA}} = \mathbf{P}\mathbf{W}_{\mathbb{R}\text{-ICA}}, \tag{4.2}$$

where $\mathbf{P} \in \mathbb{R}^{D_e \times D_e}$ *is a permutation matrix to be determined and* $\mathbf{W}_{\mathbb{R}\text{-ISA}}$ *is the* $\mathbb{R}$*-ISA demixing matrix.*

Note 3 (Connection to JBD) *It is intriguing that if* (4.1) *is satisfied then this simple decomposition principle provides the* global *minimum of the*

$$J(\mathbf{W}) := I\left(\mathbf{y}^1, \ldots, \mathbf{y}^M\right) \quad (\mathbf{W} \in \mathcal{O}^{D_e}, \mathbf{y} = \left[\mathbf{y}^1; \ldots; \mathbf{y}^M\right] = \mathbf{W}\mathbf{x}, \mathbf{y}^m \in \mathbb{R}^{d_m}) \tag{4.3}$$

ISA cost. In the literature on JBD [12] have put forth a similar conjecture *recently. According to this conjecture, for quadratic cost function, if Jacobi optimization is applied, the joint block diagonalization of the matrices can be found by the optimization of permutations following the joint diagonalization of the matrices. ISA solutions formulated within the JBD framework [170, 171, 172, 3] make efficient use of this idea in practice. [172] could justify this approach for* local *minimum points.*

Note 4 (Consequences in $\mathbb{R}$-ISA Cost Optimization) *There are several possibilities to optimize* $\mathbb{R}$*-ISA cost functions:*

1. *Without ICA preprocessing, optimization problems concern either the* Stiefel manifold *[67, 122, 144, 149] or the* flag manifold *[135]. According to our experiences, these gradient based optimization methods may be stuck in poor local minima. The problem can be reduced somewhat by smart initialization procedures [23].*

2. *According to the* $\mathbb{R}$*-ISA Separation Theorem, it may be sufficient to search for optimal permutation of the* $\mathbb{R}$*-ICA components provided by* $\mathbb{R}$*-ICA preprocessing.*

   - *Our experiences show that greedy permutation search is often sufficient for the estimation of the* $\mathbb{R}$*-ISA subspaces.*[2]

[2] Applying greedy permutation search strategy: two coordinates of different subspaces are exchanged provided that this change improves cost function $J_{\mathbb{R}\text{-ISA}}$.

- *However, it is easy to generate examples in which this is not true [145]. In such cases, global permutation search methods of higher computational burden may become necessary. We apply such a global technique, the so-called Cross-Entropy [153] to the solution of the* $\mathbb{R}$*-ISA problem in Section 5.1.*
- *Since the* $\mathbb{R}$*-ISA Separation Theorem transforms the* $\mathbb{R}$*-ISA problem into clustering, non-combinatorial approximation of* $\mathbb{R}$*-ISA, and* $\mathbb{R}$*-ISA-reducible problems will become possible (Section 5.3).*

The contribution of the present thesis to the ISA Separation Theorem is the following:

- it gives sufficient conditions for (4.1), and
- provides a complex counterpart of the Separation Theorem, the $\mathbb{C}$-ISA Separation Theorem (Section 4.4.2).

The question of which types of sources satisfy the $\mathbb{R}$-ISA Separation Theorem is open. Equation (4.1) provides only a *sufficient* condition. Below, we list a few different sources types of $\mathbf{e}^m$ that satisfy (4.1):

Theorem 5 ([9]; $\mathbb{R}$-ISA Separation Theorem–Sufficient Conditions) *The sufficient condition* (4.1) *of the* $\mathbb{R}$*-ISA Separation Theorem holds for:*

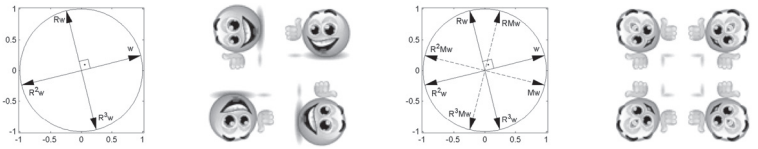
- *spherically symmetric (regarded as spherical from now on) variables [70]. The distribution of such variables is invariant to orthogonal transformations.*[3]
- *Moreover, in case of 2-dimensional components (*$d_m = 2$*) invariance to* 90° *rotation suffices. Under this condition, density function* f *of component* $\mathbf{u} = \mathbf{e}^m$ *is subject to the following invariance*

$$f(u_1, u_2) = f(-u_2, u_1) = f(-u_1, -u_2) = f(u_2, -u_1) \quad \left(\forall \mathbf{u} \in \mathbb{R}^2\right). \tag{4.4}$$

A special case of this requirement is invariance to permutation and sign changes

$$f(\pm u_1, \pm u_2) = f(\pm u_2, \pm u_1). \tag{4.5}$$

[3] In the $\mathbb{R}$-ISA task the non-degenerate affine transformations of spherical variables, the so-called *elliptical* variables, do not provide valuable generalizations due to the ambiguities of the $\mathbb{R}$-ISA task.



(a) (b) (c) (d)

Figure 4.2: Illustration: Density functions (for variables $\mathbf{e}^m$) invariant to 90° rotation *or* permutation and sign changes. (a) and (c): density function f takes identical values at the arrowheads. Matrix $\mathbf{R}$ and matrix $\mathbf{M}$ are 90° counter-clockwise rotation and reflection to axis x, respectively. (b) and (d): examples for (a) and (c), respectively.

> *In other words, there exists a function* $g : \mathbb{R}^2 \to \mathbb{R}$*, which is symmetric in its variables and* $f(\mathbf{u}) = g(|u_1|, |u_2|)$*. Special cases within this family are distributions*
>
> $$f(\mathbf{u}) = h\left(\sum_i |u_i|^p\right) \quad (p > 0), \tag{4.6}$$
>
> *which are constant over the spheres of* $L^p_{\mathbb{R}}$*-space. They are called* $L^p_{\mathbb{R}}$*-spherical variables which, for* $p = 2$*, corresponds to spherical variables.*

[See Proof IV on page 74.]

In fact, the aforementiond source component types fulfill a stronger condition than that of (4.1). This condition is called $\mathbb{R}$-w-EPI condition, where EPI is shorthand for the *entropy power inequality* [58]:

$$e^{2H\left(\sum_{i=1}^d w_i u_i\right)} \geq \sum_{i=1}^d e^{2H(w_i u_i)}, \forall \mathbf{w} \in \mathcal{S}^{d_m}_{\mathbb{R}}. \tag{4.7}$$

The $\mathbb{R}$-w-EPI condition is valid (beyond the examples of Theorem 5) for certain weakly dependent variables: [164] has determined sufficient conditions when EPI holds.[4] If the EPI property is satisfied on unit sphere $\mathcal{S}^{d_m}_{\mathbb{R}}$, then the $\mathbb{R}$-ISA Separation Theorem holds. The enumerated sufficient conditions are summarized in Fig. 4.3.

[4] The constraint of $d_m = 2$ may be generalized to higher dimensions. We are not aware of such generalizations.

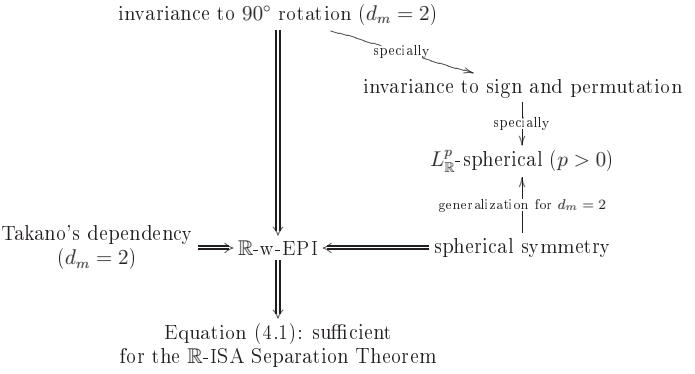


Figure 4.3: Sufficient conditions for the $\mathbb{R}$-ISA Separation Theorem.

4.2.2 Reduction of $\mathbb{R}$-uMA-IPA to $\mathbb{R}$-ISA

Below, (i) we describe the Temporal Concatenation (TCC) and Linear Prediction Approximation (LPA) separation principles of the $\mathbb{R}$-uMA-IPA task and (ii) derive their causal FIR filtering interpretations.

The Temporal Concatenation Technique (TCC)

The TCC technique for the $\mathbb{R}$-uMA-IPA task is formulated as follows:

Theorem 6 ([9]; $\mathbb{R}$-uMA-IPA via TCC) *Let L' be such that*

$$D_x L' \geq D_e(L_e + L') \tag{4.8}$$

is fulfilled.[5] *Let $x_m(t)$ denote the m^{th} coordinate of observation $\mathbf{x}(t)$. Then we end up with an $\mathbf{X}(t) = \mathbf{A}\mathbf{E}(t)$ $\mathbb{R}$-ISA task with an $\mathbf{A} \in \mathbb{R}^{D_x L' \times D_e(L_e+L')}$ ($\mathbf{H}[z]$ dependent) Toeplitz matrix upon applying temporal concatenation of depth $L_e + L'$ and L' on the sources and the observations, respectively, that is $\mathbf{E}^m(t) := [\mathbf{e}^m(t); \ldots; \mathbf{e}^m(t -$*

[5]Such L' exists due to the undercomplete assumption $D_x > D_e : L' \geq \left\lceil \frac{D_e L_e}{D_x - D_e} \right\rceil$.

$(L_e + L') + 1)]$, $\mathbf{E}(t) := [\mathbf{E}^1(t); \ldots; \mathbf{E}^M(t)]$, $\mathbf{X}^m(t) := [x_m(t); \ldots; x_m(t - L' + 1)]$, $\mathbf{X}(t) := [\mathbf{X}^1(t); \ldots; \mathbf{X}^{D_x}(t)]$. *[See Proof V on page 75.]*

The statement extends the result of [77] concerning BSD ($d_m = 1$), to the $d_m \geq 1$ case. Making use of this principle the $\mathbb{R}$-uMA-IPA problem can be reduced to $\mathbb{R}$-ISA by applying temporal concatenation. Choosing the minimal value for L', the dimension of the obtained $\mathbb{R}$-ISA task is

$$D_{\mathbb{R}\text{-ISA}} := D_e(L_e + L') = D_e\left(L_e + \left\lceil \frac{D_e L_e}{D_x - D_e} \right\rceil\right). \tag{4.9}$$

Taking into account the ambiguities of the $\mathbb{R}$-ISA task (see page 11), the original $\mathbf{e}^m$ components will occur $L_e + L^{'}$ times and up to orthogonal transformations. As a result, in the ideal case, our estimations are as follows

$$\hat{\mathbf{e}}_k^m := \mathbf{V}_k^m \mathbf{e}^m \in \mathbb{R}^{d_m}, \tag{4.10}$$

where $k = 1, \ldots, K(:= L_e + L')$, $\mathbf{V}_k^m \in \mathcal{O}^{d_m}$. The pseudocode of the TCC based solution of the $\mathbb{R}$-uMA-IPA problem can be found in Table B.1.

The Linear Prediction Technique (LPA)

The $\mathbb{R}$-ISA problem associated to $\mathbb{R}$-uMA-IPA by the TCC approach can easily become 'high dimensional' (see Eq. 4.9). This dimensionality problem can be alleviated by the LPA approach. The LPA technique is compressed into the following statement, which (i) is the extension of [71] to multidimensional $\mathbf{e}^m$ components ($d_m \geq 1$), (ii) can be extended to the $\mathbb{R}$-uARMA-IPA problem (see Theorem 8):

Theorem 7 ([10]; $\mathbb{R}$-uMA-IPA via LPA) *In the $\mathbb{R}$-uMA-IPA task, observation process $\mathbf{x}(t)$ is autoregressive and its innovation $\tilde{\mathbf{x}}(t) := \mathbf{x}(t) - E[\mathbf{x}(t)|\mathbf{x}(t-1), \mathbf{x}(t-2), \ldots]$ is $\mathbf{H}_0\mathbf{e}(t)$, where $E[\cdot|\cdot]$ denotes the conditional expectation value. Consequently, there is a polynomial matrix $\mathbf{W}_{\mathrm{AR}}^{\mathrm{LPA}}[z] \in \mathbb{R}[z]^{D_x \times D_x}$ such that $\mathbf{W}_{\mathrm{AR}}^{\mathrm{LPA}}[z]\mathbf{x} = \mathbf{H}_0\mathbf{e}$.* *[See Proof VI on page 76.]*

Thus, the AR fit of $\mathbf{x}(t)$ can be used for the estimation of $\mathbf{H}_0\mathbf{e}(t)$. This innovation corresponds to the observation of an undercomplete $\mathbb{R}$-ISA model[6], which can be reduced to a complete $\mathbb{R}$-ISA model using PCA (see Note II). Finally, the solution can be completed by any $\mathbb{R}$-ISA procedure. Taking these steps together, one can introduce polynomial matrix $\hat{\mathbf{W}}^{\mathrm{LPA}}[z] := \hat{\mathbf{W}}^{\mathrm{LPA}}_{\mathrm{ISA}}\hat{\mathbf{W}}^{\mathrm{LPA}}_{\mathrm{PCA}}\hat{\mathbf{W}}^{\mathrm{LPA}}_{\mathrm{AR}}[z] \in \mathbb{R}[z]^{D_e \times D_x}$ and claim that the LPA estimation gives rise to source estimation $\hat{\mathbf{e}} = \hat{\mathbf{W}}^{\mathrm{LPA}}[z]\mathbf{x}$, and consequently, LPA uses causal filter $\hat{\mathbf{W}}^{\mathrm{LPA}}[z]$ for FIR filtering the observation $\mathbf{x}$. The pseudocode of the LPA based solution of the $\mathbb{R}$-uMA-IPA problem can be found in Table B.2.

Connection Between TCC and LPA

Using Theorem 6, the TCC technique estimates components $\mathbf{e}^m$ of hidden source $\mathbf{e}$ as the linear transformation of the temporally concatenated form of observation $\mathbf{x}$. Thus, components $\mathbf{e}^m$ are causal FIR filtered versions of $\mathbf{x}$ similar to the LPA method (see Proof VII on page 77). The difference is in the method of the estimation:

- The LPA technique applies Gaussian approximation in the AR estimation (for $\mathbf{H}_0\mathbf{e}$) and then rotates the solution to independent directions.
- By contrast, the TCC method—via the temporal concatenation—directly optimizes the FIR filter system $\mathbf{W}^{\mathrm{TCC}}[z] = \{\mathbf{W}^{\mathrm{TCC}}_{m,k}[z]\}$ to make the estimated components $\hat{\mathbf{e}}^m_k$ $(m = 1, \ldots, M; k = 1, \ldots, K)$ independent.

Although, according to our experiments, the LPA technique is better suited for large scale problems, the TCC approach may have its own advantages (see Section 6.3.3).

4.2.3 Reduction of $\mathbb{R}$-uARMA-IPA to $\mathbb{R}$-ISA via LPA

The LPA based $\mathbb{R}$-uMA-IPA separation principle (Theorem 7) also holds for $\mathbb{R}$-uARMA-IPA problems (for the case of $d_m = 1$, see [75]):

Theorem 8 ([2]; $\mathbb{R}$-uARMA-IPA via LPA) *In the $\mathbb{R}$-uARMA-IPA task, observation process $\mathbf{x}(t)$ is autoregressive and its innovation $\tilde{\mathbf{x}}(t) := \mathbf{x}(t) - E[\mathbf{x}(t)|\mathbf{x}(t -$*

[6] Assumptions made for $\mathbf{H}[z]$ (and $\mathbf{A}$) in the $\mathbb{R}$-u(AR)MA-IPA task implies that $(\mathbf{A})\mathbf{H}_0$ is of full column rank and thus the resulting $\mathbb{R}$-ISA task is well-defined.

$1), \mathbf{x}(t-2), \ldots]$ *is* $\mathbf{H}_0\mathbf{e}(t)$*, where* $E[\cdot|\cdot]$ *denotes the conditional expectation value. Consequently, there is a polynomial matrix* $\mathbf{W}_{\mathrm{AR}}^{\mathrm{LPA}}[z] \in \mathbb{R}[z]^{D_x \times D_x}$ *such that* $\mathbf{W}_{\mathrm{AR}}^{\mathrm{LPA}}[z]\mathbf{x} = \mathbf{A}\mathbf{H}_0\mathbf{e}$.[6] *[See Proof VI on page 76.]*

The pseudocode of the LPA based solution of the $\mathbb{R}$-uARMA-IPA problem can be found in Table B.2.

4.2.4 Reduction of $\mathbb{R}$-uARIMA-IPA to $\mathbb{R}$-uARMA-IPA via Temporal Differentiating

According to Theorem 8, the solution of $\mathbb{R}$-uARMA-IPA can be transcribed to $\mathbb{R}$-ISA by the LPA technique. Our non-stationary ARIMA extension (the ARIMA-IPA problem [2]) can be reduced to $\mathbb{R}$-uARMA-IPA by temporal differentiating. Namely, let us note that differentiating the observation $\mathbf{x}$ of the $\mathbb{R}$-uARIMA-IPA task in Eq. (3.2) in r^{th} order, and making use of the relation $z\mathbf{x} = \mathbf{A}(z\mathbf{s})$, the following holds:

$$\mathbf{F}[z]\left(\nabla^r[z]\mathbf{s}\right) = \mathbf{H}[z]\mathbf{e}, \tag{4.11}$$

$$\nabla^r[z]\mathbf{x} = \mathbf{A}\left(\nabla^r[z]\mathbf{s}\right). \tag{4.12}$$

That is taking $\nabla^r[z]\mathbf{x}$ as observations, the problem at hand becomes equivalent to an $\mathbb{R}$-uARMA-IPA task. Neither the undercompleteness, nor the real-valued property of the problem have been used in this decomposition step.

4.3 Real Post Nonlinear Models

Now, we present a separation principle for the i.i.d, PNL-ISA problem [11] of the real post nonlinear models, which is—according to our *numerical experiments*—a viable way of solution up to the PNL-uARMA-IPA task.

Theorem 1 implies that by using an appropriate transformation $\mathbf{g}$ acting on each coordinate separately, the $\mathbf{d}$-independence of the estimation

$$\hat{\mathbf{e}} = \mathbf{W}\mathbf{g}(\mathbf{x}) \tag{4.13}$$

solves the PNL-ISA task. Thus, the solution of the PNL-ISA task can be carried out

in 2 steps:

1. Estimate $\mathbf{g}$: according to the d-dependent central limit theorem [143], term $\mathbf{Ae}$ can be considered as an approximately gaussian variable, so $\mathbf{g}$ can be approximated as a 'gaussianization' transformation (see [183, 160] for $d_m = 1$).

2. Estimate $\mathbf{W}$: apply a linear $\mathbb{R}$-ISA method on the result of the 'gaussianization' transformation.

4.4 Complex Linear Models

As $\mathbb{C} \cong \mathbb{R}^2$, one may think at first sight that complex-valued models *do not differ* from a real-valued model of double dimension:

- On the one hand, this is *true*: there are certain $\mathbb{C}$-ICA *methods*, which have already been rewritten to real ones (see, e.g., [43] and references therein). In fact, all linear complex *problems* of Section 3.1.2 can be transformed to real problems (Section 4.4.1). Thus, *any* real method can be applied for their solution—provided that, there is *at least one Gaussian* among the associated hidden components.

- on the other hand, this is *false*: an exciting result for the $\mathbb{C}$-ICA problem ($\forall d_m = 1$) is that the mixture of certain Gaussian variables can also be demixed [68]—which is in contrast to the real non-Gaussian constraint. In the complex case, for $d_m \geq 1$ ($\mathbb{C}$-ISA) we are not aware of any similar separability result. Thus, our $\mathbb{C}$-ISA Separation Theorem presented in Section 4.4.2 can be considered as a pioneering and *preliminary* result. The theorem implies—alike the $\mathbb{R}$-ISA separation principle—that the $\mathbb{C}$-ISA problem can be solved by (i) $\mathbb{C}$-ICA (assuming *one-dimensional* hidden independent components), (ii) followed by the grouping of the estimated $\mathbb{C}$-ICA coordinates by their statistical dependence.

4.4.1 Reduction of $\mathbb{C}$-uARIMA-IPA to $\mathbb{R}$-uARIMA-IPA

Here we reduce the tasks of Fig. 3.1, which have complex variables to real problems. In particular, we reduce the $\mathbb{C}$-uARIMA-IPA problem to the $\mathbb{R}$-uARIMA-IPA task.

Consider the mappings

$$\varphi_v : \mathbb{C}^L \ni \mathbf{u} \mapsto \mathbf{u} \otimes \begin{bmatrix} \Re(\cdot) \\ \Im(\cdot) \end{bmatrix} \in \mathbb{R}^{2L}, \tag{4.14}$$

$$\varphi_M : \mathbb{C}^{L_1 \times L_2} \ni \mathbf{M} \mapsto \mathbf{M} \otimes \begin{bmatrix} \Re(\cdot) & -\Im(\cdot) \\ \Im(\cdot) & \Re(\cdot) \end{bmatrix} \in \mathbb{R}^{2L_1 \times 2L_2}, \tag{4.15}$$

where $\otimes$ is the Kronecker product, $\Re$ stands for the real part, $\Im$ for the imaginary part, subscript 'v' ('M') for vector (matrix). Using these notations, our statement is as follows:

Theorem 9 ([4]; $\mathbb{C}$-uARIMA-IPA to $\mathbb{R}$-uARIMA-IPA) *The $\mathbb{C}$-uARIMA-IPA problem with parameters*

$$(\mathbf{x}, \mathbf{s}, \mathbf{e}, \mathbf{F}[z], \mathbf{H}[z], \mathbf{A}, L_e, L_s, r, \mathbf{d}, D_x, D_s, D_e) \tag{4.16}$$

can be realized as a $\mathbb{R}$-uARIMA-IPA task with parameters

$$(\varphi_v(\mathbf{x}), \varphi_v(\mathbf{s}), \varphi_v(\mathbf{e}), \varphi_M(\mathbf{F}[z]), \varphi_M(\mathbf{H}[z]), \varphi_M(\mathbf{A}), L_e, L_s, r, 2\mathbf{d}, 2D_x, 2D_s, 2D_e). \tag{4.17}$$

Thus, the solution of the original problem can be carried out by any $\mathbb{R}$-uARIMA-IPA technique, provided that there is at least one Gaussian *variable among the associated components $\varphi_v(\mathbf{e}^m) \in \mathbb{R}^{2d_m}$. [See Proof VIII on page 77.]*

4.4.2 The $\mathbb{C}$-ISA Separation Theorem

Now, we present our separation principle for the $\mathbb{C}$-ISA problem, which implies that the solution of the $\mathbb{C}$-ISA task can be formulated as finding the optimal permutation, grouping of the $\mathbb{C}$-ICA elements:

Theorem 10 ([7]; $\mathbb{C}$-ISA Separation Theorem) *Presume that the $\mathbf{e}^m$ sources of the $\mathbb{C}$-ISA model satisfy condition*

$$H\left(\sum_{i=1}^{L} w_i u_i\right) \geq \sum_{i=1}^{L} |w_i|^2 H(u_i) \quad \forall \mathbf{w} \in S_{\mathbb{C}}^{d_m}, \tag{4.18}$$

$\mathbb{C}$-spherical symmetry $\Longrightarrow$ $\mathbb{C}$-w-EPI $\Longrightarrow$ Equation (4.18): sufficient for the $\mathbb{C}$-ISA Separation Theorem

Figure 4.4: Sufficient conditions for the $\mathbb{C}$-ISA Separation Theorem.

and that $J_{\mathbb{C}\text{-ICA}}(\mathbf{W}) = \sum_{m=1}^{M}\sum_{i=1}^{d_m} H\left(y_i^m\right)$, $(\mathbf{W} \in \mathcal{U}^{D_e})$, *i.e., the* $\mathbb{C}$*-ICA cost function has minimum. Here,* $S_{\mathbb{C}}^{d_m}$ *denotes the* d_m*-dimensional complex unit sphere, that is* $S_{\mathbb{C}}^{d_m} := \left\{\mathbf{w} = [w_1; \ldots; w_{d_m}] \in \mathbb{C}^{d_m} : \sum_{i=1}^{d_m} |w_i|^2 = 1\right\}$. *Then it is sufficient to search for the minimum of the* $\mathbb{C}$*-ISA task (*$\mathbf{W}_{\mathbb{C}\text{-ISA}}$*) as a permutation of the solution of the* $\mathbb{C}$*-ICA task (*$\mathbf{W}_{\mathbb{C}\text{-ICA}}$*). That is, it is sufficient to search for the* $\mathbb{C}$*-ISA demixing matrix in the form*

$$\mathbf{W}_{\mathbb{C}\text{-ISA}} = \mathbf{P}\mathbf{W}_{\mathbb{C}\text{-ICA}}, \tag{4.19}$$

where $\mathbf{P}\left(\in \mathbb{R}^{D_e \times D_e}\right)$ *is the permutation matrix to be determined. [See Proof IX on page 78.]*

We proved the following sufficient conditions for the $\mathbb{C}$-ISA Separation Theorem:

Theorem 11 ([7]; $\mathbb{C}$-ISA Separation Theorem–Sufficient Conditions) *The sufficient condition* (4.18) *(formulated for hidden source* components*) of the* $\mathbb{C}$*-ISA Separation Theorem is fulfilled by:*

- complex spherical *variables [114]. The distribution of such variables is invariant to unitary transformations.*
- *variables satisfying* $\mathbb{C}$*-w-EPI.*

[See Proof XVIII on page 84. The relation of these sufficient conditions is depicted in Fig. 4.4.]

To summarize achievements presented in this chapter, the following can be stated: $\mathbb{K}$-ARIMA-IPA problems can be solved in several ways by transforming and reducing the problem to simpler setups as shown in Fig. 4.1.

Chapter 5

Specific Algorithms for Independent Process Analysis

In Chapter 4, we delineated separation principles for the $\mathbb{K}$-ARIMA-IPA problem family. By applying these principles, the resulting subproblems can now be addressed by several different methods. The goal of this chapter is to derive some efficient algorithms for these specific subproblems. Namely,

- According to the ISA Separation Theorem (see Theorem 2 and Theorem 10), under certain conditions, the solution of the ISA task can be carried out by finding the optimal permutation of the estimated ICA elements grouping them to statistically dependent subspaces. For the permutation search problem a global technique, the Cross-Entropy (CE) procedure [153] is adapted in Section 5.1.

- We develop an ISA algorithm in Section 5.2 that (i) builds upon joint decorrelation for a set of functions (hence the name JFD, Joint $\mathcal{F}$-Decorrelation), (ii) can be related to kernel based techniques, (iii) can be interpreted as a self-adjusting, self-organizing neural network solution, (iv) is a first step towards large scale problems. Our numerical examples extend to a few 100-dimensional ISA tasks. Such dimensions can easily appear, e.g., in the TCC based solution of the $\mathbb{R}$-uMA-IPA problem (see Section 4.2.2).

- Former, existing ISA methods assume that the dimensions of the different subspaces (d_m) are known in advance. In Section 5.3 we present techniques that can

estimate the unknown dimensions as well. These techniques provide solutions for ISA(-reducible) problems (see Fig. 4.1).

5.1 The Cross-Entropy Method for $\mathbb{R}$-ISA

In the present section, the CE procedure is tailored to the ISA problem: it is applied for grouping the estimated ICA elements. The CE method has been found efficient for combinatorial optimization problems [153]. The CE technique operates as a two step procedure: First, the problem is converted to a stochastic problem and then the following two-phases are iterated (for detailed description, see the above reference):

1. Generate $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_N \in \mathcal{X}$ samples from a distribution family parameterized by a θ parameter and choose the *elite* of the samples. The elite is the best $\rho\%$ of the samples according to the cost function J.

2. Modify the sample generation procedure (θ) according to the elite samples. In practice, smoothing, i.e., $\theta^{new} = \beta \cdot \theta^{proposed} + (1-\beta) \cdot \theta^{old}$ is employed in the update of θ, where $\beta \in [0,1]$.

This technique can be applied in our search for permutation matrix $\mathbf{P}$. Our method is similar to the CE solution suggested for the Travelling Salesman Problem (TSP) (see [153]). In the TSP problem, a permutation of cities is searched for. The objective is to minimize the cost of the travel. We are also searching for a permutation, but now the travel cost is replaced by $J_{\mathbb{R}\text{-ISA}}(\mathbf{W})$. Thus, in our case, $\mathcal{X} = S_{D_e}$ and $\tilde{\mathbf{x}}$ is an element of this permutation group. Further, the CE cost equals to

$$J(\mathbf{P}_{\tilde{\mathbf{x}}}) = J_{\mathbb{R}\text{-ISA}}(\mathbf{P}_{\tilde{\mathbf{x}}}\mathbf{W}_{\mathrm{ICA}}), \tag{5.1}$$

where $\mathbf{P}_{\tilde{\mathbf{x}}}$ denotes the permutation matrix associated to $\tilde{\mathbf{x}}$. Thus, optimization concerns permutations in $\mathcal{X}$. θ contains transition probabilities $i \to j$ ($1 \le i, j \le D_e$), called *node transition* parametrization [153]. The above iteration is stopped if there is no change in the cost (in the last L steps), or the change in parameter θ is negligibly small (smaller then ϵ). Our approach has been illustrated in [6] using the $J_{\mathbb{R}\text{-ISA}}(\mathbf{W}) = \sum_{m=1}^{M} H(\mathbf{y}^m)$ multidimensional entropy based ISA cost function, where $\mathbf{y} = [\mathbf{y}^1; \ldots; \mathbf{y}^M] = \mathbf{W}\mathbf{x}$ ($\mathbf{y}^m \in \mathbb{R}^{d_m}$) denotes the estimated hidden source.

5.2 The Joint $\mathcal{F}$-Decorrelation Technique (JFD)

Now, we present an ISA method, which builds on joint decorrelation on a function set [3]. Components $\mathbf{e}^m$ are estimated by a demixing network, which aims to 'decorrelate' (see below) the $\mathbf{y}^m \in \mathbb{R}^{d_m}$ parts of the $\mathbb{R}^{D_e} \ni \mathbf{y}(t) = [\mathbf{y}^1(t); \ldots; \mathbf{y}^M(t)]$ output of the network. The demixing network is chosen to be a neural network executing the mapping

$$\mathbf{x} \mapsto \mathcal{N}(\mathbf{x}, \mathbf{\Theta}) \tag{5.2}$$

with parameter $\mathbf{\Theta}$. We describe possible network architectures ($\mathcal{N}$) in Section 5.2.1, Section 5.2.2 is about decorrelation using a single function setting the stage for our Joint $\mathcal{F}$-Decorrelation Technique (JFD) ISA method detailed in Section 5.2.3.

5.2.1 Demixing Network Candidates

Choosing an RNN (Recurrent Neural Network) with feedforward ($\mathbf{F}$) and recurrent ($\mathbf{R}$) connections then the network assumes the form

$$\dot{\mathbf{y}}(\tau) = -\mathbf{y}(\tau) + \mathbf{F}\mathbf{x}(t) - \mathbf{R}\mathbf{y}(\tau) \tag{5.3}$$

and thus, upon relaxation it solves the

$$\mathbf{y}(t) = (\mathbf{I} + \mathbf{R})^{-1}\mathbf{F}\mathbf{x}(t) = \mathcal{N}(\mathbf{x}(t); \mathbf{F}, \mathbf{R}) \tag{5.4}$$

input-output mapping [128, 17]. Another natural choice is a network with feedforward connections $\mathbf{W}$ that executes mapping

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) = \mathcal{N}(\mathbf{x}(t); \mathbf{W}). \tag{5.5}$$

5.2.2 Cost Function of the Demixing Network

The network estimates hidden sources $\mathbf{e}^m$ by non-linear ($\mathbf{f}$) decorrelation of $\mathbf{y}^m$s, components of network output $\mathbf{y}$. Formally: Let us denote the empirical $\mathbf{f}$-covariance matrix of $\mathbf{y}(t)$ and $\mathbf{y}^m(t)$ for function $\mathbf{f} = [\mathbf{f}^1; \ldots; \mathbf{f}^M]$ over $[1, T]$ (T: sample number)

by

$$\mathbf{\Sigma}(\mathbf{f},T) = \widehat{cov}\left(\mathbf{f}(\mathbf{y}),\mathbf{f}(\mathbf{y})\right), \tag{5.6}$$

$$\mathbf{\Sigma}^{i,j}(\mathbf{f},T) = \widehat{cov}\left(\mathbf{f}^i\left(\mathbf{y}^i\right),\mathbf{f}^j\left(\mathbf{y}^j\right)\right), \tag{5.7}$$

respectively, where $i,j = 1,\ldots,M$. Then minimization of the following non-negative cost function (in $\mathbf{\Theta}$)

$$J(\mathbf{\Theta};\mathbf{f},T) := -\frac{1}{2}\log\left\{\frac{\det[\mathbf{\Sigma}(\mathbf{f},T)]}{\prod_{m=1}^{M}\det[\mathbf{\Sigma}^{m,m}(\mathbf{f},T)]}\right\} \tag{5.8}$$

gives rise to *pairwise*[1] $\mathbf{f}$-uncorrelatedness:

Theorem 12 (Pre-JFD Cost) *For the demixing carried out by the network minimizing cost function* (5.8)*, the following statements are equivalent:*

i) $\mathbf{f}$*-uncorrelatedness:* $\mathbf{\Sigma}^{i,j}(\mathbf{f},T) = 0 \quad (\forall i \neq j)$.

ii) $J(\cdot;\mathbf{f},T)$ *is minimal:* $J(\mathbf{\Theta},\mathbf{f},T) = 0$.

[See Proof XIX on page 84.]

Note 13

1. *For the special case,* $\mathbf{\Theta} = (\mathbf{F},\mathbf{R})$, $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ *and* $\forall d_m = 1$*, see [128].*

2. *Cost function* J *of* (5.8) *is attractive as its gradient can easily be computed. This gradient for the case of an RNN architecture [see Eq.* (5.4)*] may give rise to self-organization [128].*

3. *The demixing defined by cost function* (5.8)*, can be related to the KGV technique [23]. This technique aims to separate the* $\mathbf{y}^m$ *independent components of* $\mathbf{y}$*, the transformed form of input* $\mathbf{x}$*. To this end, KGV estimates mutual information* $I(\mathbf{y}^1,\ldots,\mathbf{y}^M)$ *in Gaussian approximation. Here, the transformation of the KGV*

[1] We note that if our observations are generated by an ISA model then—unlike in the ICA task when $d_m = 1$, $\forall m$—pairwise independence is *not* equivalent to mutual independence [55, 145]. Nonetheless, according to our numerical experiences it is an efficient approximation in many situations.

technique is realized by the neural network parameterized with variable $\mathbf{\Theta}$ *and by the function* $\mathbf{f}$.

4. *We note that KGV is related to the KC method [78], which makes use of the supremum of one-dimensional covariances as a measure of independence. Our approximation may also be improved by minimizing* $J(\mathbf{\Theta}; \mathbf{f}, T)$ *on* $\mathcal{F}(\ni \mathbf{f})$, *i.e., on a set of functions.*

5.2.3 $\mathbb{R}$-ISA by JFD

From now on, we are working with the linear feedforward neural network architecture [see, Eq. (5.5)], and attempt to recover the hidden sources $\mathbf{e}^m$ by pairwise decorrelation of the components $\mathbf{y}^m$ of the output of the network using function manifold $\mathcal{F}$ ($\mathcal{F}$: see the last note). Thus by making use of the $\mathbb{R}$-ISA Separation Theorem and Theorem 12 our cost function is

$$J_{\mathrm{JFD}}(\mathbf{P}; \mathcal{F}, T) := \sum_{\mathbf{f} \in \mathcal{F}} \|\mathbf{M} \circ \mathbf{\Sigma}(\mathbf{f}, T, \mathbf{P})\|^2 \to \min_{\mathbf{P}}. \tag{5.9}$$

Here:

- (i) $\mathcal{F}$ denotes a set of functions, each function $\mathbb{R}^{D_e} \mapsto \mathbb{R}^{D_e}$, and each function acts on each coordinate separately, (ii) $\circ$ denotes pointwise multiplication (Hadamard product), (iii) $\mathbf{M}$ is a mask according to the subspaces [$\mathbf{M} = \mathbf{E} - blockdiag(\mathbf{E}_1, \ldots, \mathbf{E}_M)$, where all elements of matrix $\mathbf{E} \in \mathbb{R}^{D_e \times D_e}$ and $\mathbf{E}_m \in \mathbb{R}^{d_m \times d_m}$ are equal to 1], (iv) $\|\cdot\|^2$ denotes the square of the Frobenius norm (sum of squares of the elements), (v) in $\mathbf{\Sigma}(\mathbf{f}, T, \mathbf{P})$, $\mathbf{y} = \mathbf{P}\hat{\mathbf{e}}_{\mathrm{ICA}}$, and (vi) $\mathbf{P}$ is the $D_e \times D_e$ permutation matrix to be determined.

- The JFD cost function, J_{JFD} can be interpreted as follows: Independence of the estimated sources $\mathbf{y}^m$ is approximated by the uncorrelatedness on the function set $\mathcal{F}$, which is equivalent to the minimization of J_{JFD} according to Theorem 12.

- To optimize J_{JFD}, one may apply greedy permutation search: 2 coordinates of different subspace are exchanged if this change lowers cost function $J_{\mathrm{JFD}}(\cdot; \mathcal{F}, T)$. The pseudocode of this JFD implementation can be found in Table B.3. Note:

Greedy search could be replaced by a *global* at a price of larger computational load, see Section 5.1.

This simple JFD based ISA approximation—with greedy permutation search—could cope with about 300-400-dimensional (D_e) ISA problems on a standard PC (see Section 6.3.1), while former existing state-of-the art ISA techniques can only handle problems of about 20 dimensions.

5.3 Estimation of the Hidden Component Dimensions in ISA-Reducible Problems

Most existing ISA, IPA methods assume that the dimensions (d_m) of the hidden components ($\mathbf{e}^m$) are known. In this section, we propose solutions not requiring this a priori knowledge. The lack of this knowledge may cause *combinatorial difficulty* in a sense that one may have to try all possible

$$D_e = d_1 + \ldots + d_M \quad (d_m > 0, M \leq D_e) \tag{5.10}$$

combinations for the subspace dimensions. The number of these possibilities is given by the so-called partition function $f(D_e)$, i.e., the number of sets of positive integers that sum up to D_e. The value of $f(D_e)$ grows quickly with the argument. Asymptotic behavior is known [82, 175]:

$$f(D_e) \sim \frac{e^{\pi\sqrt{2D_e/3}}}{4D_e\sqrt{3}}, \quad D_e \to \infty. \tag{5.11}$$

Fortunately, making use of the ISA Separation Theorem efficient *non-combinatorial approximations* can be constructed. The basic idea behind these approximations is to 'cluster by the dependencies between the coordinates of the estimated sources'. In addition, this non-combinatorial method can also be applied to all problems that can be transformed into an ISA task based on the separation principles presented in Chapter 4.

Note 14 *So far, we addressed the question of approximating the dimensions of the* individual *hidden components* $\mathbf{e}^m$*, provided that the total dimension* D_e *is known. It*

may also occur that the value of D_e is not given beforehand. Let us suppose, that this is the case and let us deal with the ISA problem. Because the observation takes the form $\mathbf{x} = \mathbf{Ae}$, the eigenvalues of the covariance matrix of the observation can reveal the value of D_e: ideally there are D_e positive and $D_x - D_e$ (almost) 0 values. Estimation is always performed based on a finite sample set, thus in practice $D_x - D_e$ of the eigenvalues are near 0, then there is a sharp transition. This train of thoughts concerned ISA. However, the same technique can be applied to the result of innovation/gaussianization [11] by the separation principles of Chapter 4. Thus, in the sequel, it is assumed that D_e is known and our task is to estimate the d_m dimensions.

My results estimating dimensionality in IPA problems [8, 1, 5, 2] with the approaches known from the literature [147, 172, 161, 24] can be summarized as follows:

AR-IPA with Block-Diagonal F: [147] ($L_s = 1$)**, [8]** ($L_s \geq 1$) Using the fact that in the AR-IPA problem [(3.5)-(3.6)] the observation is also AR, and that the basis transformation rule implies

$$\mathbf{x}(t) = \sum_{i=1}^{L_s} \mathbf{A}\mathbf{F}_i^s\mathbf{A}^{-1}\mathbf{x}(t-i) + \mathbf{Ae}(t) \tag{5.12}$$

(or shortly, $\mathbf{x} = \mathbf{F}^x[z]\mathbf{x} + \mathbf{Ae}$,) the predictive matrix[2] of the hidden source ($\mathbf{F}^s[z] = \sum_{i=1}^{L_s} \mathbf{F}_i^s z^i$) takes the form

$$\mathbf{F}^s[z] = \mathbf{W}\mathbf{F}^x[z]\mathbf{W}^{-1} = \mathbf{W}\mathbf{F}^x[z]\mathbf{W}^T. \tag{5.13}$$

Thus, making use of the ISA Separation Theorem, polynomial matrix

$$\mathbf{W}_{\mathrm{ICA}}\mathbf{F}^x[z]\mathbf{W}_{\mathrm{ICA}}^T \tag{5.14}$$

– apart from (possible) permutations – is equal to the block-diagonal predictive matrix $\mathbf{F}^s[z]$ of the source $\mathbf{s}$. It then follows that connected groups of the coordinates of the hidden source can be recovered by collecting the elements that belong to the same block in $\mathbf{F}^s[z]$. In practice, the estimation of

[2]Note: as the predictive matrices of $\mathbf{s}$ and $\mathbf{x}$ also appear in the paragraph, they get a disticnt superscript ($\mathbf{F}^s[z], \mathbf{F}^x[z]$) to avoid confusion.

matrix $\mathbf{F}^s[z]$ (i.e., matrix $\hat{\mathbf{F}}^s[z]$), is only nearly block-diagonal (apart from permutation). Thus, we say that two coordinates i and j are $\hat{\mathbf{F}}^s[z]$-'connected' if $\max(|\hat{F}^s_{ij}|, |\hat{F}^s_{ji}|) > \epsilon$, where, $|\hat{F}^s_{ij}| = \sum_{k=1}^{L_s} |\hat{F}^s_{k,ij}|$, $\hat{F}^s_{k,ij}$ denotes the $(i,j)^{th}$ coordinate of the k^{th} matrix from $\hat{\mathbf{F}}^s[z]$, and, in the ideal case, $\epsilon = 0$. Then we can group the $\hat{\mathbf{F}}^s[z]$-'connected' coordinates into separate subspaces using the following algorithm:

1. Choose an arbitrary coordinate i ($1 \leq i \leq D_s$) and group all $j \neq i$ coordinates, which are $\hat{\mathbf{F}}^s[z]$-'connected' with it.
2. Choose an arbitrary and not yet grouped coordinate. Find its connected coordinates recursively. Group them together.
3. Continue until all components are grouped.

This *gathering procedure* is fast. In the worst case, it is quadratic in D_s.

ISA, Cumulant Based Matrices: Similar considerations can be applied in the ISA problem by replacing $[|\hat{F}^s_{ij}|]_{i,j=1,\ldots,D_e} \in \mathbb{R}^{D_e \times D_e}$, for example, with cumulant based matrices [172].

AR-IPA, Iterative Solution: The above-mentioned AR-IPA solution can be implemented iteratively [1]. Here, the gathering procedure is unchanged, but the estimation of $\mathbf{F}^s[z]$ and $\mathbf{W}$ is carried out iteratively. The optimization is initilized by a random $\hat{\mathbf{F}}^s[z]$ and performs iteratively the following two steps:

1. perform ICA on the estimated innovation of the observation (on $\mathbf{x} - \hat{\mathbf{W}}^{-1}\hat{\mathbf{F}}^s[z]\hat{\mathbf{W}}\mathbf{x}$) to update $\hat{\mathbf{W}}$,
2. improve AR fit on the estimated source $\hat{\mathbf{s}} = \hat{\mathbf{W}}\mathbf{x}$, to get the new $\hat{\mathbf{F}}^s[z]$ estimation.

According to our numerical experiments, in batch mode this technique converges to the optimum in 3-4 iterations.

ISA, ISA-reducible AR-IPA: We present a JFD based technique for the estimation of the ISA subspace dimensions in [5]. This method

- in contrast to the approach of [8] ($L_s \geq 1$): is using the following matrix

$$\hat{\mathbf{C}}^{\mathcal{F}} := \sum_{\mathbf{f} \in \mathcal{F}} |\hat{\mathbf{C}}_{\mathbf{f}}|, \tag{5.15}$$

  instead of the estimated predictive matrix $\hat{\mathbf{F}}^e[z]$ (ideally the latter is $\mathbf{0}$ because of the i.i.d. assumption of ISA). Here, $\hat{\mathbf{C}}_{\mathbf{f}}$ is the empirical correlation matrix of the ICA element for function $\mathbf{f} \in \mathcal{F}$. In other words, $\hat{\mathbf{C}}_{\mathbf{f}} = \hat{corr}[\mathbf{f}(\hat{\mathbf{e}}_{\mathrm{ICA}}), \mathbf{f}(\hat{\mathbf{e}}_{\mathrm{ICA}})]$; $|\cdot|$ denotes absolute values for all coordinates.[3]

- similarly to the approach of [8] ($L_s \geq 1$) : is a $\hat{\mathbf{F}}^e[z]$-'connected' type solution by the

$$\hat{\mathbf{F}}^e[z] := \sum_{\mathbf{f} \in \mathcal{F}} \hat{\mathbf{C}}_{\mathbf{f}} z^{[\mathbf{f}]} \tag{5.16}$$

  correspondence. Here, $[\mathbf{f}]$ denotes the index (in some order) of the function $\mathbf{f}$ coming from a finite element function set $\mathcal{F}$.

This approach also provides solution for the AR-IPA problem by applying it to the estimated innovation delivered by the AR fit (see Fig. 4.1; [5]). Thus, we can alternatively (see item 'AR-IPA with Block-Diagonal $\mathbf{F}$') work exploiting the dependencies of the ICA-ized innovation ($\{\hat{\mathbf{C}}_{\mathbf{f}}\}_{\mathbf{f} \in \mathcal{F}}$), instead of the predictive matrix of the innovation of the estimated hidden source $\mathbf{s}$ ($\hat{\mathbf{F}}^s[z]$).

Weaknesses of the above threshold based methods include

- the uncertainty in choosing the threshold ε, and
- the fact that the methods are sensitive to the threshold. Nevertheless, this sensitivity can be alleviated to some extent by increasing the number of samples [5, 1, 8], see Section 6.3.2.

More robust solutions can be designed if dependencies, for example, the mutual information among the coordinates, are used to construct an adjacency matrix and apply a clustering method for this matrix. Then, one might use, for example, hierarchical [161] or tree-structured clustering methods [24] to solve the ISA task.

[3] Note: we computed correlations for matrices $\mathbf{C}_f$ (instead of covariances; see Section 5.2) because they are normalized.

The possibility of estimating the unknown dimensions can be extended to ISA-reducible problems by the separation principles of Chapter 4. This approach has been illustrated in [2] by applying a variant of the NCut algorithm [182]: the grouping revealed, interesting 'face-component'-type subspaces.

Chapter 6

Illustrations

This chapter is devoted to the illustration of our separation principles and algorithms (Chapter 4, Chapter 5). Test databases are introduced in Section 6.1. The quality of the solutions will be measured by the normalized Amari-error, the Amari-index (Section 6.2). Numerical results are presented in Section 6.3.

6.1 Databases

We conducted experiments on 15 databases, with their different parameterization, to assess the efficiency, robustness and limits of our techniques. These databases can be split into 4 categories:[1]

Category$_1$**:** Sources from this category satisfy (one of the) sufficient conditions of the ISA Separation Theorem (Fig. 4.3). Databases:

1. *d-spherical* (d: scalable, $M \leq 3$) [3, 1],
2. *mosaic* ($d = 2$, $M \leq 4$) [8].

Category$_2$**:** Sources from this category satisfy the conditions of the studied model, but we don't know whether they are from *Category*$_1$, or not. Databases:

1. *3D-geom* ($d = 3$, $M \leq 6$) [6, 5, 10, 11, 9, 8],
2. *ABC* ($d = 2$, $M \leq 50$) [3, 5, 10, 11, 9],

[1] Databases are ordered alphabetically in each category.

3. *all-k-independent* ($d = k + 1$: scalable, M: scalable) [6, 9],
4. *celebrities* ($d = 2$, $M \leq 10$) [5, 10, 11, 9],
5. *d-geom* (d: scalable, $M \leq 4$) [4],
6. *numbers* ($d = 2$, $M \leq 10$) [6],
7. *smiley* ($d = 2$, $M \leq 6$) [6],
8. *tale* ($d = 2$, $M \leq 6$) [1].

Category$_3$**:** Sources from this category are outside the conditions of the studied models, but we 'know' the values of d_m and M. Databases:

1. *Beatles* ($d = 2$, $M = 2$) [10, 9],
2. *IFS*[2] ($d = 2$, $M \leq 9$) [11],
3. *Led Zeppelin* ($d = 2$, $M \leq 4$) [8],
4. *Lorenz* ($d = 3$, $M \leq 3$) [8].

Category$_4$**:** Source from this category is outside the conditions of the studied models, and we do not know the values of d_m and M.

1. *FaceGen* [2]: In this database[3] face images were compressed by PCA to 60 dimensions ($D_e = 60$), and we searched for $M = 4$ ISA subspaces.

In the present thesis, our goal is to *summarize* our numerical experiments—the interested reader is referred to the cited papers for further results. Databases used for the illustration are now detailed:

6.1.1 The 3D-geom, the ABC and the Celebrities Database

The first 3 databases are illustrated in Fig. 6.1. In the *3D-geom* test $\mathbf{e}^m$s were random variables uniformly distributed on 3-dimensional geometric forms ($d = 3$). We chose 6 different components ($M = 6$) and, as a result, the dimension of the hidden source $\mathbf{e}$ is $D_e = 18$. In the *ABC* database, hidden sources $\mathbf{e}^m$ were uniform distributions defined by 2-dimensional images ($d = 2$) of the English+Greek alphabet. M can be at most

[2]IFS stands for Iterated Function System.
[3]`http://www.facegen.com/modeller.htm`

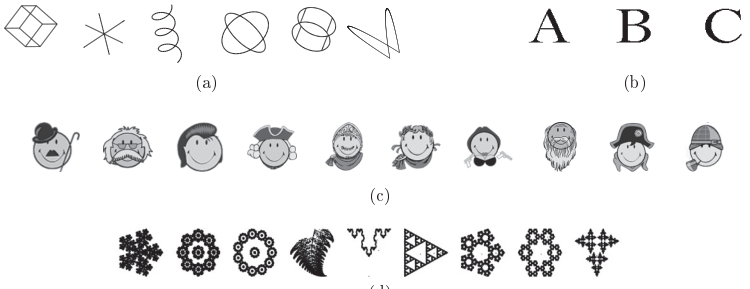


Figure 6.1: Illustration of the *3D-geom*, *ABC*, *celebrities* and *IFS* databases. (a): database *3D-geom*, 6 pieces of 3-dimensional components ($M = 6$, $d = 3$). Hidden sources are uniformly distributed variables on 3-dimensional geometric objects. (b): database *ABC*. Here, the hidden sources $\mathbf{e}^m$ are uniformly distributed on images ($d = 2$) of letters. (c): database *celebrities*. Density functions of the hidden sources are proportional to the pixel intensities of the 2-dimensional images ($d = 2$). Number of hidden components: $M = 10$. (d): dataset *IFS*. Here components are self-similar structures generated from iterated function systems: $d = 2$, $M = 9$.

$26 + 24 = 50$. The *celebrities* test has 2-dimensional source components generated from cartoons of celebrities ($d = 2$, $M = 10$).[4] Sources $\mathbf{e}^m$ were generated by sampling 2-dimensional coordinates proportional to the corresponding pixel intensities. In other words, 2-dimensional images of celebrities were considered as density functions. $M = 10$ was chosen.

6.1.2 The all-k-independent Database

The d-dimensional hidden components $\mathbf{u} := \mathbf{e}^m$ were created as follows: coordinates $u_i(t)$ $(i = 1, \ldots, k)$ were uniform random variables on the set {0,...,k-1}, whereas u_{k+1} was set to $mod(u_1 + \ldots + u_k, k)$. In this construction, every k-element subset of $\{u_1, \ldots, u_{k+1}\}$ is made of independent variables. This database is called the *all-k-independent* problem [145, 6], the dimension of the components $d = k + 1$ can be varied.

[4] See http://www.smileyworld.com.

(a) (b)

Figure 6.2: Illustration of the *d-geom* and *d-spherical* databases with $d = 3$ and $d = 2$, respectively. (a): database *d-geom*, 4 pieces of d-dimensional variables distributed uniformly on geometrical forms. (b): stochastic representation of the 3 hidden d-dimensional sources of the *d-spherical* database ($M = 3$). Left: ρ is uniform on $[0, 1]$, center: ρ is exponential with parameter $\mu = 1$, right: ρ is lognormal with parameters $\mu = 0$, $\sigma = 1$, respectively.

6.1.3 The d-geom Database

In the *d-geom* dataset $\mathbf{e}^m$s were random variables uniformly distributed on d-dimensional geometric forms. Geometrical forms were chosen as follows. We used: (i) the surface of the unit ball, (ii) the straight lines that connect the opposing corners of the unit cube, (iii) the broken line between $d+1$ points $\mathbf{0} \to \mathbf{e}_1 \to \mathbf{e}_1+\mathbf{e}_2 \to \ldots \to \mathbf{e}_1+\ldots+\mathbf{e}_d$ (where $\mathbf{e}_i$ is the i canonical basis vector in $\mathbb{R}^d$, i.e., all of its coordinates are zero except the i, which is 1), and (iv) the skeleton of the unit square. Thus, the number of components M was equal to 4, and the dimension of the components (d) is scalable. For illustration, see Fig 6.2(a).

6.1.4 The d-spherical Database

Here, hidden sources $\mathbf{e}^m$ were spherically symmetric random variables that have representation of the form $\mathbf{v} \stackrel{\text{distr}}{=} \rho\mathbf{u}^{(d)}$, where $\mathbf{u}^{(d)}$ is uniformly distributed on the d-dimensional unit sphere, and ρ is a non-negative scalar random variable independent of $\mathbf{u}^{(d)}$ ($\stackrel{\text{distr}}{=}$ denotes equality in distribution). This *d-spherical* database: (i) can be scaled in dimension d, (ii) satisfies conditions of the $\mathbb{R}$-ISA Separation Theorem, and (iii) can be defined by ρ. Our choices for ρ are shown in Fig. 6.2(b). The number of the components M was 3.

6.1.5 The Beatles Database

Our *Beatles* test is a non-i.i.d. example. Here, hidden sources are stereo Beatles songs.[5] 8 kHz sampled portions of two songs (A Hard Day's Night, Can't Buy Me Love) made the hidden $\mathbf{e}^m$s. Thus, the dimension of the components d was 2 and the number of components M was 2.

6.1.6 The IFS Database

Dataset *IFS* is a non-i.i.d. example: components $\mathbf{e}^m$ are realizations of IFS based 2-dimensional ($d = 2$) self-similar structures. For all m we have chosen the following triple: $(\{\mathbf{h}_k\}_{k=1,\ldots,K}, \mathbf{p} = (p_1, \ldots, p_K), \mathbf{v}_1\}$, where (i) $\mathbf{h}_k : \mathbb{R}^2 \to \mathbb{R}^2$ are affine transformations in the form $\mathbf{h}_k(\mathbf{z}) = \mathbf{C}_k\mathbf{z} + \mathbf{d}_k$ ($\mathbf{C}_k \in \mathbb{R}^{2\times 2}, \mathbf{d}_k \in \mathbb{R}^2$), (ii) $\mathbf{p}$ is a distribution over the indices $\{1, \ldots, K\}$ ($\sum_{k=1}^K p_k = 1, p_k \geq 0$), and (iii) for the initial value we chose $\mathbf{v}_1 := (\frac{1}{2}, \frac{1}{2})$. We generated T samples in the following way: (i) $\mathbf{v}_1$ is given ($t = 1$), (ii) an index $k(t) \in \{1, \ldots, K\}$ was drawn according to the distribution $\mathbf{p}$ and the next sample is generated as $\mathbf{v}_{t+1} := \mathbf{h}_{k(t)}(\mathbf{v}_t)$. The resulting series $\{\mathbf{v}_1, \ldots, \mathbf{v}_T\}$ was taken as a hidden source component $\mathbf{e}^m$ and this way we generated 9 components ($M = 9$) to make the *IFS* dataset [see Fig. 6.1(d)].

6.2 Performance Measure, the Amari-index

First, let us suppose, that we are dealing with the $\mathbb{R}$-ISA task, all components are d-dimensional ($d = d_m$, $\forall m$), and the performance of a $\mathbb{R}$-ISA estimation ($\hat{\mathbf{W}}_{\text{ISA}}$) is to be measured. The optimal estimation provides matrix $\mathbf{G} := \hat{\mathbf{W}}_{\text{ISA}}\mathbf{A} \in \mathbb{R}^{D_e \times D_e}$, a block-permutation matrix made of $d \times d$ sized blocks. This block-permutation property can be measured by the *Amari-index*. Namely, let matrix $\mathbf{G}$ be decomposed into $d \times d$ blocks: $\mathbf{G} = [\mathbf{G}^{ij}]_{i,j=1,\ldots,M}$. Let g^{ij} denote the sum of the absolute values of the elements of matrix $\mathbf{G}^{ij} \in \mathbb{R}^{d\times d}$. Then the normalized version of the Amari-error [18] adapted to the ISA task [170, 171] is defined as [6]:

$$r(\mathbf{G}) := \frac{1}{2M(M-1)}\left[\sum_{i=1}^{M}\left(\frac{\sum_{j=1}^{M} g^{ij}}{\max_j g^{ij}} - 1\right) + \sum_{j=1}^{M}\left(\frac{\sum_{i=1}^{M} g^{ij}}{\max_i g^{ij}} - 1\right)\right]. \tag{6.1}$$

[5] See http://rock.mididb.com/beatles/.

We refer to the normalized Amari-error as the Amari-index. One can see that $0 \leq r(\mathbf{G}) \leq 1$ for any matrix $\mathbf{G}$, and $r(\mathbf{G}) = 0$ if and only if $\mathbf{G}$ is a block-permutation matrix with $d \times d$ sized blocks. $r(\mathbf{G}) = 1$ is in the worst case, i.e, when all the elements of $\mathbf{G}$ are equal in absolute value.

In case of the $\mathbb{R}$-uARIMA-IPA problem, using the results of Section 4, ideally, the product of matrix $\mathbf{AH}_0$ and the matrices provided by PCA, ISA, i.e., $\mathbf{G} := (\hat{\mathbf{W}}_{\text{ISA}}\hat{\mathbf{W}}_{\text{PCA}})\mathbf{AH}_0 \in \mathbb{R}^{D_e \times D_e}$ is a block-permutation matrix made of $d \times d$ blocks. Thus, one can measure the block-permutation property of this $\mathbf{G}$ matrix by the Amari-index. In case of $\mathbb{C}$-uARIMA-IPA, ideally $\mathbf{G} := (\hat{\mathbf{W}}_{\text{ISA}}\hat{\mathbf{W}}_{\text{PCA}})\varphi_M(\mathbf{A})\varphi_M(\mathbf{H}_0) \in \mathbb{R}^{2D_e \times 2D_e}$ is a block-permutation matrix made of $2d \times 2d$ blocks, which can be similarly measured by the Amari-index.

6.3 Simulations

In what follows, our numerical experiences are summarized, in 'Topic–Illustration' pairs: first the addressed issues, questions are posed, then numerical illustrations are provided.

6.3.1 ISA Algorithms (CE, JFD, $\mathbb{C}$-ISA Separation Theorem)

The ISA problem can be reduced by the ISA Separation Theorem (Section 2 and Section 4.4.2) to finding the optimal permutation of the ICA elements.

Topic

1. In certain cases, global permutation search methods of higher computational burden may become necessary. The CE technique (presented in Section 5.1, and see [6]) offers an efficient approach in this situation.

2. Independence of the estimated source components may be approximated by the joint decorrelation over a function set $\mathcal{F}$. This simple approximation, clustering the ICA elements by the JFD based cost function (see Section 5.2 and [3]), offers a first step towards large scale ISA problems: with greedy search, one may tackle few hundred dimensional tasks.

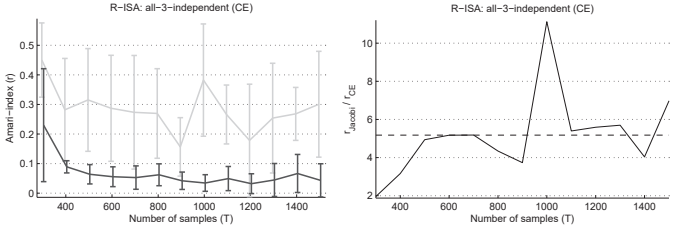


Figure 6.3: Illustration of the CE based ISA algorithm. Left: average ± standard deviation of the Amari-index as a function of the sample number, gray–Jacobi method, black–CE method. Right: relative precision of the estimations, dashed–average over the different sample numbers.

3. In the complex case, according to our experiments, the $\mathbb{C}$-ISA Separation Theorem based solution can provide more precise estimations than its ' $\mathbb{C}$-ISA $\rightarrow\mathbb{R}$-ISA' based counterpart.

Illustration

1. Here, the task was to solve an $\mathbb{R}$-AR-IPA problem. The hidden AR processes were driven by $M = 5$ pieces of *all-3-independent* sources ($d = 3 + 1 = 4$)[6], and then mixed by a random orthogal matrix $\mathbf{A}$. After the identification of the AR observation process, the innovation was analyzed by $\mathbb{R}$-ISA. The 'global' CE approach was compared by the 'greedy' Jacobi algorithm of [148], which applies Jacobi rotations for any pairs of the elements received after $\mathbb{R}$-ICA preprocessing. $\mathbb{R}$-ICA preprocessing was performed by [92]. Numerical values of the CE parameters were chosen as $\rho = 0.05$, $\beta = 0.4$, $L = 7$, $\epsilon = 0.005$. Sample number T was incremented by 100 between 300 and 1500, and we averaged the results, the Amari-indices of 10 computer runs to measure the performance of the algorithms. The precision of the procedures is shown in Fig. 6.3 as a function of the sample number. One can see, that the CE method is superior for all sample

[6] According to our experiments, the *all-k-independent* construction is a quite challenging dataset for ISA methods.

Table 6.1: Amari-index for the JFD method and database *d-spherical*, for different d values: average $\pm$ standard deviation. Number of samples: $T = 30,000$.

$d = 20$	$d = 30$	$d = 40$	$d = 50$	$d = 60$
1.40% ($\pm$0.03)	1.71% ($\pm$0.03)	1.99% ($\pm$0.03)	2.23% ($\pm$0.03)	2.44% ($\pm$0.03)
$d = 70$	$d = 80$	$d = 90$	$d = 100$	$d = 110$
2.65% ($\pm$0.03)	2.85% ($\pm$0.03)	3.03% ($\pm$0.04)	3.19% ($\pm$0.02)	3.37% ($\pm$0.03)

numbers [Fig. 6.3(a)]. For $T = 1,500$ samples the Jacobi method has precision ($100r\%\pm$standard deviation) of 30.05%($\pm$17.90). The same precision for the CE technique is 4.31($\pm$5.61), on the $300 - 1,500$ sample interval this means a $1.96 - 5.18 - 11.12$-times (min-mean-max) improvement. The relative precision of the two methods is depicted in Fig. 6.3(b).

2. In our *d-spherical* $\mathbb{R}$-ISA illustration (Fig. 6.4, Table 6.1) scaling properties of the approximation were studied by changing the value of d between 20 and 110 [i.e., the number of subspaces (M) was fixed, but the dimension of the subspaces was increased.] For each parameters $[(T, d)]$ ten experiments were averaged. Qualities of the solutions were measured by the Amari-index. Sample number of observations $\mathbf{x}(t)$ changed $1,000 \leq T \leq 30,000$, mixing matrix $\mathbf{A}$ was chosen randomly from the orthogonal group, manifold $\mathcal{F}$ was $\mathcal{F} := \{\mathbf{z} \mapsto \cos(\mathbf{z}), \mathbf{z} \mapsto \cos(2\mathbf{z})\}$ (functions operated on coordinates separately). We have chosen FastICA [92] for the $\mathbb{R}$-ICA module (see Table B.3). Precision of our method is shown Fig. 6.4 as a function of sample number and source dimension (d) (for details, see Table 6.1). The figure demonstrates that the algorithm was able to uncover the hidden components with high precision, and the Amari-index decreases according to power law $r(T) \propto T^{-c}$ ($c > 0$). In our numerical simulations, the number of sweeps before the iteration of the permutation optimization stopped (see Table B.3) varied between 2 and 6.

   We note that

   - the performance of our algorithms presented here often (almost always) show this power law behavior.
   - A direct application of the JFD method is the TCC based solution of the

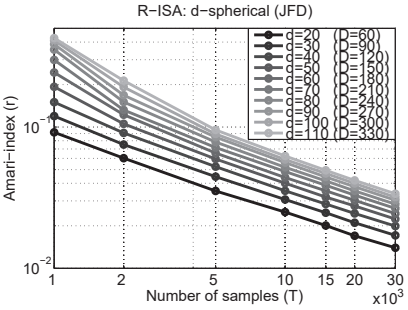


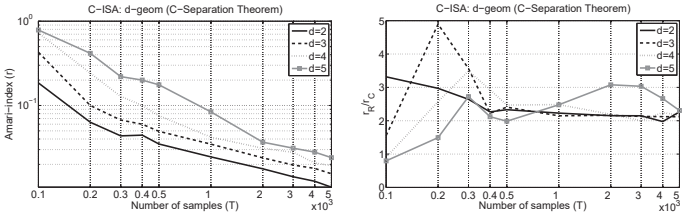
Figure 6.4: Precision of the JFD method for database *d-spherical*: Amari-index as a function of the number of samples on log-log scale for different dimensional (d) subspaces. Task dimension: $D = D_e = D_x$. Errors are approximately linear, so they scale according to power law, like $r(T) \propto T^{-c}$ ($c > 0$). For numerical values, see Table 6.1.

$\mathbb{R}$-uMA-IPA problem (Section 4.2.2), which requires solving 'high dimensional' ISA tasks, see Section 6.3.3 for an illustration.

- one may easily attack larger ISA problems, by applying large(r)-scale ICA methods. In sum, the ICA preprocessing represents the only computational bottleneck here.

3. In this illustration the *d-geom* dataset is used to drive a $\mathbb{C}$-ISA problem. The hidden sources $\mathbf{e}^m \in \mathbb{C}^d$ were defined in $\mathbb{R}^{2d}$ by the '*2d-geom*' construction and φ_v^{-1} derived images were taken as $\mathbf{e}^m \in \mathbb{C}^d$. The mixing matrix $\mathbf{A}$ [see, Eq. (3.2)] was drawn randomly from the unitary group.

   We studied the performance of the $\mathbb{C}$-ISA algorithms by (i) changing the value of d (dimension in complex sense) between 2 and 5 [i.e., the number of subspaces (M) was fixed, but the dimension of the subspaces was increased.] and (ii) varying the sample size T between 100 and 1,500. For each parameters $[(T, d)]$ ten experiments were averaged. Qualities of the solutions were measured by the Amari-index (in $\mathbb{R}^{2D\times 2D}$ measuring the block-permutation property with $2d\times 2d$ size blocks). To perform the ICA preprocessing step, the fastICA method [92], and its complex counterpart [33] was applied. The clustering of the estimated



(a) (b)

Figure 6.5: Estimation error of the $\mathbb{C}$-ISA Separation Theorem based $\mathbb{C}$-ISA method. (a) Amari-index as a function of the sample number and dimension of the components on log-log scale. (b) Quotients of the Amari-indices of the $\mathbb{C}$-ISA Separation Theorem and the '$\mathbb{C}$-ISA $\rightarrow\mathbb{R}$-ISA' methods: for quotient value $q > 1$, the former method is q times more precise.

ICA elements was based on the k-nearest neighbours based entropy cost [146].

The precision of the $\mathbb{C}$-ISA Separation Theorem based algorithm is shown in Fig. 6.5 as a function of the sample number and source component dimension (d). The figure demonstrates that although both $\mathbb{C}$-ISA algorithms were able to uncover the hidden components with high precision [, and the Amari-index decreases according to power law $r(T) \propto T^{-c}$ $(c > 0)$] the complex based solution is more precise ($2-3$ times on the average) than the real technique. Improvements were $2.43(\pm 0.42)$, $2.53(\pm 0.97)$, $2.35(\pm 0.67)$, $2.27(\pm 0.71)$ times for $d = 2, 3, 4, 5$ on the average ($\pm$standard deviation)–the average was computed over the number of samples and the 10 random simulations.

6.3.2 Estimation of the Unknown Dimensions, Gaussianization

Topic

1. We use the PNL-ISA problem (see Section 4.3 and [11]) to illustrate:

   (a) the goodness of the gaussianization as a function of the dimension of the hidden source.

(b) the possibility to estimate the dimension of the hidden source, i.e., D_e, when it is not given beforehand (see Note 14),

2. Then, we illustrate to what extent non-combinatorial approximations make it possible to estimate the dimensions of the hidden source components (see Section 5.3). Below we address the $\mathbb{R}$-ISA scenario [5]–one can get similar results on $\mathbb{R}$-ISA-reducible problems [1, 8]. Experimental studies concerned the following problems:

   (a) The quality of the gathering procedure depends on the threshold parameter ε. We studied the estimation error, the Amari-index, as a function of sample number. The ε values were preset to reasonably good values.

   (b) We studied the optimal domain for the ε values. We looked for the dynamic range, i.e., the ratio of the highest and lowest 'good ε values'.

Illustration

1. (a) For the first illustration, the ABC dataset is used: we studied the Amari-index as a function of the sample number and the number of the hidden components M was chosen as $2, 3, 4, 10, 20, 50$. As for small M the Gaussian assumption on $\mathbf{Ae}$ is less likely, we expect to see deterioration of the goodness of the estimation in this range. The gaussianization was based on the ranks of samples [183], the solution of the $\mathbb{R}$-ISA task was performed by the JFD method [3], $\mathbb{R}$-ICA preprocessing was carried out by fastICA [92]. The sample size T was chosen between $1,000$ and $100,000$. The goodness of the estimation was measured by the Amari-index. For a given sample size T the goodness of 50 random runs $(\mathbf{A}, \mathbf{e}, \mathbf{f})$ were averaged. $\mathbf{A}$ was a random orthogonal matrix. The nonlinear functions f_i have been generated as

$$f_i(z) = c_i[a_i z + \tanh(b_i z)] + d_i, \tag{6.2}$$

that is they are mixtures of random, scaled and translated *id* and tanh functions. Here $a_i \in [0, 0.5]$, $b_i \in [0, 5]$, $d_i \in [0, 2]$ are random variables of uniform distribution, c_i take ± 1 values with probability $\frac{1}{2}, \frac{1}{2}$.

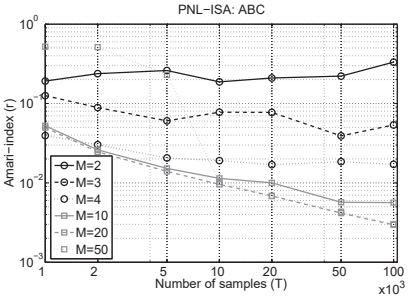


Figure 6.6: PNL-ISA: average Amari-index as a function of the sample size, on log-log scale for dataset ABC with different number of components (M).

3D-geom	celebrities	IFS
0.29%(±0.05)	0.40%(±0.03)	0.46%(±0.06)

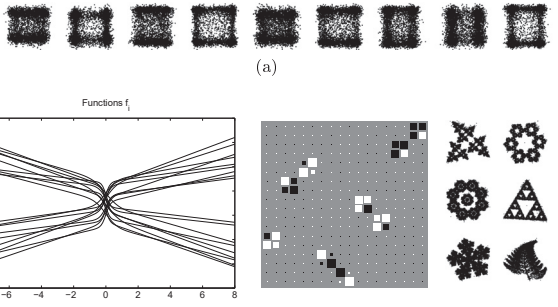
Table 6.2: PNL-ISA: Amari-index for *3D-geom*, *celebrities* and *IFS* datasets—average ± standard deviation. Sample size: $T = 100,000$.

In Fig. 6.6 the power law decline of the estimation error is presented for increasing D_e. For $M \geq 3$, (that is when $D_e \geq 6$) the PNL-ISA solution is already efficient. It can also be seen that for the case of $M = 50$ number of components (that is the dimension of the task is $D_e = 100$) we need at least $10,000$ samples to get a more reliable estimation, while for $3 \leq M < 50$ $2,000-5,000$ samples are sufficient. The exact error are shown in Table 6.3. We experienced similar power law behavior of the estimation error in the referred work for the *3D-geom* ($d = 3$, $M = 6$), *celebrities* ($d = 2$, $M = 10$) and *IFS* ($d = 2$, $M = 9$) databases. For $T = 100,000$, the errors for these datasets are given in Table 6.2 and Fig. 6.7 shows an illustration for the *IFS* test.

(b) Next we illustrate to what extent we can guess the overall dimension D_e of the hidden source $\mathbf{e}$ when it is not given beforehand (see Note 14). The dimension of the PNL-ISA observation $\mathbf{x}$ was set to $D_x = 2D_e$ (D_e of course is not available for the algorithm). The mixing matrix $\mathbf{A} \in$

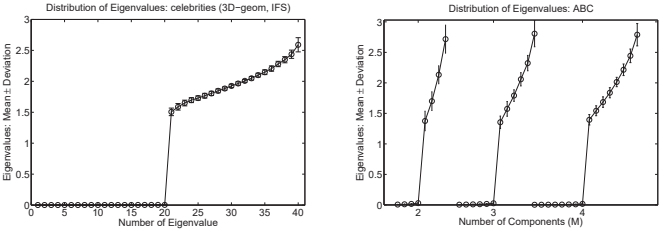
$M = 2$	$M = 3$	$M = 4$	$M = 10$	$M = 20$	$M = 50$
33.20%(±39.42)	5.37%(±8.82)	1.71%(±0.52)	0.56%(±0.50)	0.30%(±0.03)	0.30%(±0.01)

Table 6.3: PNL-ISA: Amari-index for dataset *ABC*, as a function of the number of components M—values shown are average ± standard deviation. Sample size: $T = 100,000$. The error as a function of sample size T is plotted in Fig. 6.6.



(a) (b) (c) (d)

Figure 6.7: Illustration of the PNL-ISA estimation on the dataset *IFS*. Sample size: $T = 100,000$. (a): the observed mixed $\mathbf{x}$ signal. (b): the nonlinear f_i functions. (c) the Hinton-diagram of $\mathbf{G}$, ideally it is a block-permutation matrix with blocks of size 2×2. (d): the estimated hidden components ($\hat{\mathbf{e}}^m$)–up to the ambiguities of the PNL-ISA problem.



(a) (b)

Figure 6.8: PNL-ISA: Estimation of the dimension of the hidden source $\mathbf{e}$. The ordered eigenvalues of the covariance matrix of the transformed signal $\mathbf{g}(\mathbf{x})$ are plotted. Results are averaged over 50 runs. (a): dataset *celebrities*; results for *3D-geom* and *IFS* are similar. (b): eigenvalues for the dataset *ABC*, at different number of components M.

$\mathbb{R}^{D_x \times D_e}$ was generated by first creating a random orthogonal matrix of size $D_x \times D_x$ then choosing its first D_e columns. Postnonlinearities were generated according to Eq. (6.2). Gaussianization has been done on the observations and then we studied the eigenvalues of the covariance matrix of the resulting transformed signal $\mathbf{g}(\mathbf{x})$: ideally there are D_e positive and $D_x - D_e$ (almost) 0 values. We show the ordered eigenvalues on dataset *celebrities* averaged over 50 runs in Fig. 6.8. It can be seen that exactly half of the eigenvalues are near 0, then there is a big leap. (For datasets *3D-geom* and *IFS* we have got similar results, data is not shown.) Figure 6.8(b) shows the results corresponding to dataset *ABC* for different number of components M: only the $M \leq 4$ cases are illustrated, but in the whole range of $2 \leq M \leq 50$ there is a sharp transition similar to the results gained for the other 3 datasets.

2. Results on databases *3D-geom* ($d = 3$, $M = 6$), *celebrities* ($d = 2$, $M = 10$), and *ABC* ($d = 2$, $M = 10$) are provided here. Our gauge to measure the quality of the results is the Amari-index that we computed by averaging over 50 random runs, i.e., random choice of quantities $\mathbf{A}$ and $\mathbf{e}$. In our simulations, sample number T of observations $\mathbf{x}(t)$ was varied between $1,000$ and $20,000$. Mixing

matrix $\mathbf{A}$ was generated randomly from the orthogonal group. The dynamic range is defined as follows: We divided interval $[0, \hat{C}^{\mathcal{F}}_{max}]$ ($\hat{C}^{\mathcal{F}}_{max} := \max_{i,j} \hat{C}^{\mathcal{F}}_{ij}$) into 200 equal parts. For different sample numbers in all databases at each division point we used the gathering procedure to group the $\mathbb{R}$-ICA elements. For each of the 50 random trials we have computed the Amari-indices separately. For the smallest Amari-index, we determined the corresponding interval of ε's, these are the 'good ε values'. Then we took the ratio of the largest and smallest ε values in this set and averaged the ratios over the 50 runs. The average is called the dynamic range.

Our results are summarized in Fig. 6.9. According to Fig. 6.9(a), there are good ε parameters for the $\hat{\mathbf{C}}^{\mathcal{F}}$-'connectedness' already for $1,000 - 2,000$ samples: our method can find the hidden components with high precision. Figure 6.9(a) also shows that by increasing the sample number the Amari-index decreases. For $20,000$ samples, the Amari-index is 0.5% for the *3D-geom*, 0.75% for the *celebrities*, and 0.75% for the *ABC* database, respectively on the average. The decline of the Amari-index follows power law manifested by straight line on log-log scale. Figure 6.9(b) demonstrates that for larger sample numbers threshold parameter ε that determines the $\hat{\mathbf{C}}^{\mathcal{F}}$-'connected' property can be chosen from a broader domain; the dynamic range grows. For the *3D-geom*, the *celebrities* and the *ABC* databases the measured dynamic ranges are 4.45, 5.09 and 2.05 for $20,000$ samples and for the different databases, respectively on the average.

Finally, we illustrate the quality and the working of our method in Fig. 6.10. The figure depicts the *3D-geom* test and we used $T = 20,000$ samples. According to this figure, the algorithm was able to uncover the hidden components up to the ambiguities of the $\mathbb{R}$-ISA task.

6.3.3 $\mathbb{R}$-uMA-IPA Alternatives

Topic

Our experiences on the TCC [9] and LPA [10] based solution of the $\mathbb{R}$-uMA-IPA problem (see Section 4.2.2) can be summarized as follows:

1. the LPA method is better than the TCC one for larger tasks, i.e., for tasks of

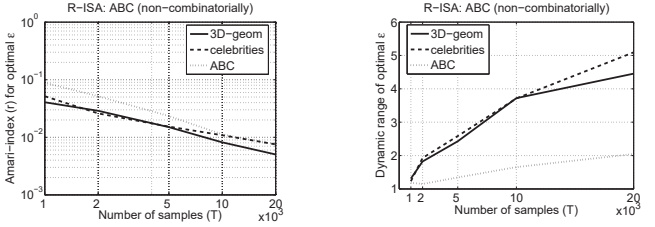


Figure 6.9: Non-combinatorial $\mathbb{R}$-ISA approximation: Amari-index on log-log scale (a) and dynamic range (b) as a function of sample number for the *3D-geom*, *celebrities*, and *ABC* databases.

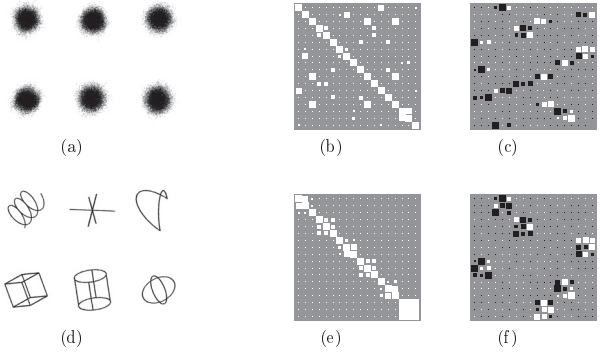


Figure 6.10: Non-combinatorial $\mathbb{R}$-ISA approximation: illustrations. (a): observed mixed signal $\mathbf{x}(t)$, (b) $\hat{\mathbf{C}}^{\mathcal{F}}$ - the sum of absolute values of the elements of the non-linear correlation matrices used for the grouping of the $\mathbb{R}$-ICA coordinates, (c): the product of the $\mathbb{R}$-ICA demixing matrix and the mixing matrix, (d): estimated components $\hat{\mathbf{e}}(t)$–up to ambiguities of the ISA problem–, based on (e): $\hat{\mathbf{C}}^{\mathcal{F}}$ after grouping, (f) product of the estimated $\mathbb{R}$-ISA demixing matrix and the mixing matrix: with high precision, it is a block-permutation matrix made of 3×3 blocks.

higher dimensions (D_e) or deeper convolutions (L_e).

2. in smaller tasks, the TCC based approximation have its own advantages. Namely, according to our experiments the AR estimation, and thus the LPA technique seems to be more sensitive to almost degenerate problems; when some of the coordinates have relatively small standard deviations.

Illustration

1. In the first illustration, the TCC and the LPA methods are compared on $\mathbb{R}$-uMA-IPA tasks. Parallely, the performance and the limits of the LPA technique are studied as a function of convolution length. We studied the $D_x = 2D_s$ case, like in [9]. Both the TCC and the LPA method reduce the $\mathbb{R}$-uMA-IPA task to $\mathbb{R}$-ISA problems and we use the Amari-index to measure and compare their performances. For all values of the parameters (sample number: T, convolution length: $L_e + 1$), we have averaged the performances upon 50 random initializations of $\mathbf{e}$ and $\mathbf{H}[z]$. The coordinates of matrices $\mathbf{H}_j$ were chosen independently from standard normal distribution. We used the Schwarz's Bayesian Criterion to determine the optimal order of the AR process. The criterion was constrained: the order Q of the estimated AR process (see Table B.2) was limited from above, the upper limit was set to twice the length of the convolution, i.e., $Q \leq 2(L_e+1)$. The AR process and the $\mathbb{R}$-ISA subtask of TCC and LPA were estimated by the method detailed in [133, 154], and by JFD [3], respectively.

   We studied the dependence of the precision versus the sample number. In the *3D-geom* ($d = 3$, $M = 6$) and *celebrities* ($d = 2$, $M = 10$) [*ABC* ($d = 2$, $M = 2$) and *Beatles* ($d = 2$, $M = 2$)] tests, the sample number T varied between $1,000$ and $100,000$ ($1,000$ and $75,000$), the length of the convolution (L_e+1) changed between 2 and 6 (2 and 31). Comparison with the TCC method and the estimations of the LPA technique are illustrated in Figs. 6.11(a)-(b) [Figs. 6.11(c)-(d)] on the *3D-geom* (*Beatles*) database. According to Fig. 6.11(a), the LPA algorithm is able to uncover the hidden components with high precisions on the *3D-geom* database. We found that the Amari-index r decreases according to power law for sample numbers $T > 2000$. According to Fig. 6.11(b) the LPA method is superior to the TCC method (i) for all sample numbers $1,000 \leq T \leq$

$100,000$, moreover (ii) LPA can provide reasonable estimates for much smaller sample numbers on the *3D-geom* database. This behavior is manifested by the initial steady increase of the quotients of the Amari indices of the TCC and LPA methods as a function of sample number followed by a sudden drop when the sample number enables reasonable TCC estimations, too. Similar results were found on the *celebrities* and the ABC databases. The LPA method resulted in $1.1-88$, $1.0-87$, $1.2-110$-times increase of precision for the *3D-geom*, *celebrities* and ABC database, respectively. For the *3D-geom* (*celebrities*, ABC) dataset the Amari-index for sample number $T = 100,000$ ($T = 100,000$, $T = 75,000$) is $0.19-0.20\%$ ($0.33-0.34\%$, $0.30-0.36\%$) with small $0.01-0.02$ (0.02, $0.11-0.15$) standard deviations.

Visual inspection of Fig. 6.11(c) shows that on the *Beatles* database the LPA method found the hidden components for sample number $T \geq 30,000$. The TCC method gave reliable solutions for sample number $T = 50,000$ or so. According to Fig. 6.11(d) the LPA method is more precise than TCC for $T \geq 30,000$. The increase in precision becomes more pronounced for larger convolution parameter L_e. Namely, for sample number $75,000$ and for $L_e = 1, 2, 5, 10, 20, 30$ the ratios of precision are 1.50, 2.24, 4.33, 4.42, 9.03, 11.13, respectively on the average. For sample number $T = 75,000$ the Amari-index stays below 1% on average $(0.4 - 0.71\%)$ and has $0.02 - 0.08$ standard deviation for the *Beatles* test.

According to our simulations, the LPA method may provide acceptable estimations for sample number $T = 20,000$ ($T = 15,000$) up to convolution length $L_e = 20$ ($L_e = 230$) for the *3D-geom* and *celebrities* (ABC and *Beatles*) datasets. Such estimations are shown in Fig. 6.12(d), Fig. 6.12(h) and Fig. 6.12(i)-(m) for the *3D-geom*, ABC and *celebrities* tests, respectively.

2. For our second illustration, the TCC and LPA $\mathbb{R}$-uMA-IPA methods were compared on 3 databases: *3D-geom* ($d = 3$, $M = 2$), ABC ($d = 2$, $M = 3$) and *celebrities* ($d = 2$, $M = 3$). Thus, the dimension of the hidden source $\mathbf{e}$ was $D_e = 6$, in all 3 tests. The almost degenerate property of hidden source $\mathbf{e}$ was modeled through an almost degenerate coordinate. That is, covariance matrix of $\mathbf{e}$ was scaled as $cov(\mathbf{e}) = diag(\varepsilon, 1, \ldots, 1)$ and we investigated the limit $(0 <)\varepsilon \to 0$. The measure of undercompleteness was $D_x = 2D_e$.

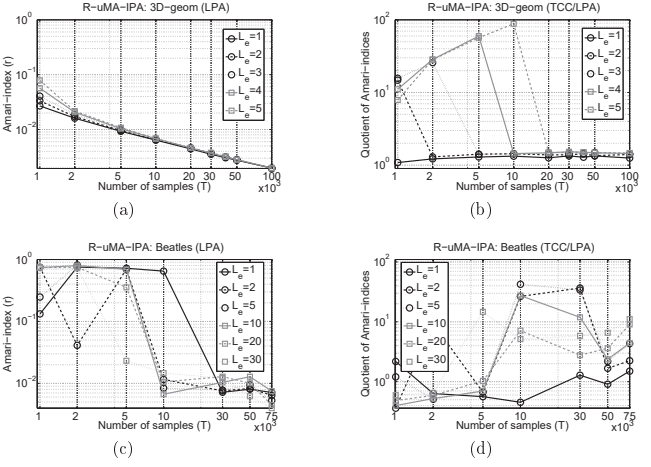


Figure 6.11: Estimation error of the LPA method and comparisons with the TCC method for the *3D-geom* and *Beatles* databases. Scales are log-log plots. Data correspond to different convolution lengths ($L_e + 1$). (a) and (c): Amari-index as a function of the sample number. (b) and (d): Quotients of the Amari-indices of the TCC and the LPA methods: for quotient value $q > 1$, the LPA method is q times more precise than the TCC method. In the *celebrities* and *ABC* tests, we found similar results as on the *3D-geom* data set.

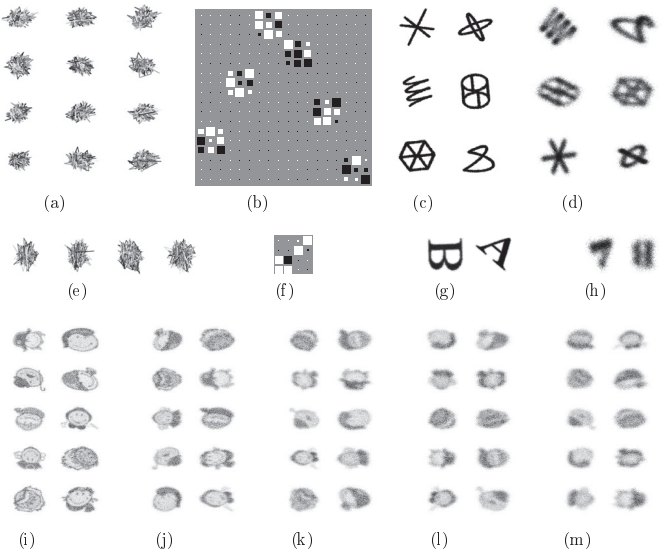
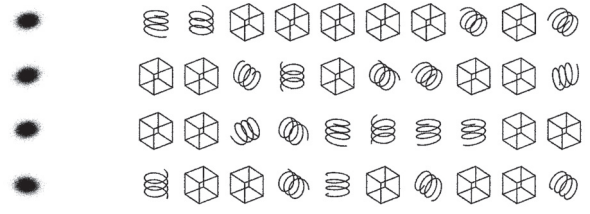


Figure 6.12: Illustration of the LPA method on the $\mathbb{R}$-uMA-IPA task for the *3D-geom* (ABC) and *celebrities* databases. (a)-(c) [(e)-(g)]: sample number $T = 100,000$ [$T = 75,000$], convolution length $L_e + 1 = 6$ [$L_e + 1 = 31$]. (a), (e): observed convolved signals $\mathbf{x}(t)$. (b) [(f)]: Hinton-diagram of $\mathbf{G}$, ideally block-permutation matrix with 3×3 [2×2] blocks. (c) [(g)]: estimated components ($\hat{\mathbf{e}}^m$), Amari-index: 0.2% [0.3%]. (d) [(h)]: estimation of hidden components ($\hat{\mathbf{e}}^m$) for sample number $T = 20,000$ [$T = 15,000$] and convolution parameter $L_e = 20$ [$L_e = 230$]. (i)-(m): sample number $T = 20,000$, dependence of estimated components ($\hat{\mathbf{e}}^m$) on the convolution parameter L_e running on values $1, 5, 10, 15, 20$.



(a) (b)

Figure 6.13: Illustration of the TCC method on the $\mathbb{R}$-uMA-IPA task for the *3D-geom* database: $\varepsilon = 10^{-3}$, convolution parameter $L_e = 10$, sample number $T = 100,000$. (a): observed convolved signals $\mathbf{x}(t)$. (b): estimated components $(\hat{\mathbf{e}}_k^m)$.

The performance of the algorithms were measured by the Amari-index: we averaged 20 random computations $(\mathbf{e}, \mathbf{H}[z])$ with fixed parameters (T, ε, L_e) to measure the quality of the estimations. The coordinates of $\mathbf{H}[z]$ were chosen independently from standard normal distribution. Sample number T varied between 500 and 100,000, convolution parameter L_e and parameter ε took values on $1, 2, 5, 10$ and $10^{-1}, 10^{-2}, 10^{-3}$, respectively.

Our results are shown in Fig. 6.14 for the *3D-geom* database for parameters $L_e = 2$ and $L_e = 10$. According to Fig. 6.14(a), the smaller the value of ε, the more advantageous the TCC method ever the LPA technique is: For $\varepsilon = 10^{-1}$, LPA is better, for $\varepsilon = 10^{-2}$ TCC becomes somewhat more precise, and for $\varepsilon = 10^{-3}$ the error of the LPA method increases with sample number, but the TCC technique gives rise to a stable estimation. These features remain true upon increasing parameter L_e—, although the necessary sample number for the TCC technique increases faster with L_e, because the $\mathbb{R}$-ISA task associated to the TCC method becomes 'high dimensional'—as it can be seen in Fig. 6.14(b). Similar results were found for the *ABC* and *celebrities* tests. Quantitative results for $L_e = 10$ are provided in Table 6.4. Figure 6.13 shows the estimation of the TCC technique for the *3D-geom* database.

Table 6.4: Amari-index in percentages for the TCC and the LPA $\mathbb{R}$-uMA-IPA methods for database *3D-geom*, *ABC* and *celebrities*, for different ε parameters, for convolution length $L_e = 10$: average $\pm$ standard deviation. Number of samples: $T = 100,000$. For other sample numbers between $500 \leq T < 100,000$ and for $L_e = 2$ see Figure 6.14.

	$\varepsilon = 10^{-1}$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-3}$
3D-geom (LPA)	0.27%($\pm$0.10%)	1.74%($\pm$0.65%)	14.25%($\pm$5.74%)
3D-geom (TCC)	0.38%($\pm$0.02%)	1.63%($\pm$0.17%)	6.69%($\pm$0.63%)
ABC (LPA)	0.51%($\pm$0.18%)	2.94%($\pm$1.10%)	9.79%($\pm$2.51%)
ABC (TCC)	0.55%($\pm$0.02%)	2.63%($\pm$0.22%)	5.05%($\pm$0.28%)
celebrities (LPA)	0.61%($\pm$0.11%)	4.06%($\pm$0.69%)	8.70%($\pm$1.55%)
celebrities (TCC)	0.57%($\pm$0.02%)	2.65%($\pm$0.16%)	5.07%($\pm$0.35%)

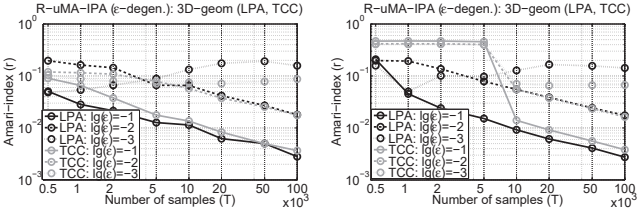


Figure 6.14: Comparisons of estimation errors of the TCC and LPA methods for the *3D-geom* database for different εs on log-log scale. (a): $L_e = 2$, (b): $L_e = 10$.

6.3.4 Complex Problems, Towards Non-Stationarity

Topic

Now the efficiency of the proposed $\mathbb{C}$-uARIMA-IPA solution is examined [4]–one may get similar results in case of the $\mathbb{R}$-uARIMA-IPA problem. We focused on 2 distinct issues:

1. How does the estimation error scale with (i) the number of samples and (ii) dimension of the hidden components?

2. We assumed that matrix polynomial $\mathbf{F}[z]$ of Eq. (3.1) is stable, that is, $\det(\mathbf{F}[z])$ has no roots within the closed complex unit circle. In the case of $r = 0$ this means that process $\mathbf{s}$ is stationary. For $r > 1$ the model describes non-stationary processes. It is expected that if the roots of $\mathbf{F}[z]$ are close to the unit circle then our estimation will deteriorate. We investigated this by generating polynomial matrix $\mathbf{F}[z]$ as follows:

$$\mathbf{F}[z] = \prod_{i=1}^{L_s} (\mathbf{I} - \lambda \mathbf{U}_i z) \quad (|\lambda| < 1, \lambda \in \mathbb{R}) \tag{6.3}$$

   Matrices $\mathbf{U}_i \in \mathbb{C}^{D_s \times D_s}$ were random unitary matrices and the $\lambda \to 1$ limit was studied.

Illustration

In our $\mathbb{C}$-uARIMA-IPA simulations:

- the *d-geom* database (d: scalable, $M = 4$) is used for illustration purposes. The hidden sources $\mathbf{e}^m \in \mathbb{C}^d$ were defined in $\mathbb{R}^{2d}$ by the '*2d-geom*' construction and φ_v^{-1} derived images were taken as $\mathbf{e}^m \in \mathbb{C}^d$.

- the measure of undercompleteness was 2 ($D_x = D_s = 2D_e$),

- the Amari-index was used to measure the precision of our method. For all values of parameters (T, L_s, r, L_e), the average performances upon 20 random initializations of $\mathbf{e}, \mathbf{H}[z], \mathbf{F}[z]$ and $\mathbf{A}$ were taken.

In economic computations, the value of r is typically ≤ 2, we investigated values between $1 \leq r \leq 3$. The coordinates of matrices $\mathbf{H}_j$ in the MA part (see Eq. (3.1)) were chosen independently and uniformly from the $\{\mathbf{v} = v_1 + iv_2 \in \mathbb{C} : -\frac{1}{2} \leq v_1, v_2 \leq \frac{1}{2}\}$ complex unit square. The mixing matrix $\mathbf{A}$ [see, Eq. (3.2)] was drawn randomly from the unitary group. Polynomial matrix $\mathbf{F}[z]$ was generated according to Eq. (6.3). The choice of λ is detailed later. The order of the AR estimation (see Fig. 4.1) was constrained from above as follows $deg(\hat{\mathbf{W}}_{\mathrm{AR}}[z]) \leq 2(L_e + 1) + L_s$ (i.e., two times the MA length + the AR length). We used the technique of [133] with the Schwarz's Bayesian Criterion to determine the optimal order of the AR process. We applied the method of [3] to solve the $\mathbb{R}$-ISA task.

1. In our first test ('small task') sample number T ranged between $2,000 \leq T \leq 30,000$ and the orders of the AR and MA processes were kept low: $L_s = 1$, $L_e = 1$ (MA order: $L_e + 1 = 2$). The order parameter r of the $\mathbb{C}$-ARIMA process was set to $r = 1, 2$ and 3 in the different computations. Sample number varied as $T = 2, 5, 10, 20, 30 \cdot 10^3$. Scaling properties of the algorithm were studied by changing the value of the dimension of the components d between 1 and 15. The value of λ was 0.9 [see, Eq. (6.3)]. Our results are summarized in Fig. 6.15(e), with an illustrative example given in Fig. 6.15(a)-(d).[7] According to Fig. 6.15(e), our method could recover the hidden components with high precision. The Amari-index $r(T)$ follows power law $r(T) \propto T^{-c}$ $(c > 0)$. The power law is manifested by straight lines on log-log scales. The slope of the lines are about the same for different d values. The actual values of the Amari-index can be found in Table 6.5 for sample number $T = 30,000$.

2. In the second test we increased parameters L_s and L_e to 5 and 10, respectively since in the 'small test' ($L_s = 1$, $L_e = 1$) we did not see relevant performance drops even for $\lambda = 0.99$. The sample number was set to $T = 20,000$. Dimension d of components $\mathbf{e}^m$ was 5. ARIMA parameter r took values on $1, 2$ and 3. Results are shown in Fig. 6.15(f). According to this figure, there is a sudden change in the performance at around $\lambda = 0.9 - 0.95$. Estimations for ARIMA parameters $r = 1, 2$ and 3 have about the same errors. We note that for $L_s = 1$

[7] The $r = 1$ case is illustrated, results are similar in the studied $r \leq 3$ domain.

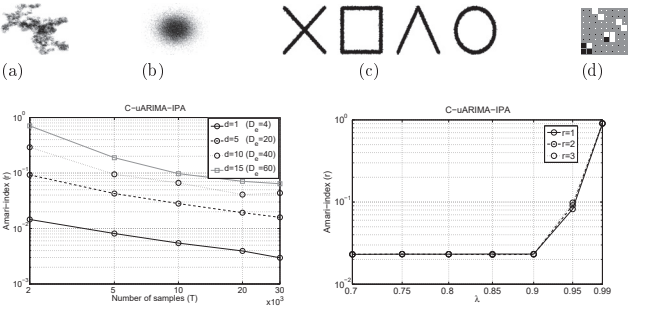


Figure 6.15: Illustration of our $\mathbb{C}$-uARIMA-IPA method. (a)-(d): AR order $L_s = 1$, MA order $L_e = 1$, order of integration $r = 1$, sample number $T = 30,000$. (a)-(b): typical 2D projection of the observed mixed $\mathbf{x}$ signal, and its r^{th}-order difference. (c): estimated components $[\varphi_v(\mathbf{e}^m)]$. (d): Hinton-diagram of $\mathbf{G}$, ideally block-permutation matrix with 2×2 blocks. (e): average Amari-index as a function of the sample size on log-log scale for different dimensional (d) components; $\lambda = 0.9$, $L_s = 1$, $L_e = 1$, $r = 1$ ($r \leq 3$). For $T = 30,000$, the exact errors are shown in Table 6.5. (f): Estimation error on log scale as a function of the magnitude of the roots of matrix polynomial $\mathbf{F}[z]$. (If $\lambda = 1$ then the roots are on the unit circle.) ARIMA parameters: $r = 1, 2$ and 3; AR order: $L_s = 5$; MA order: $L_e = 10$.

$d = 1$	$d = 5$	$d = 10$	$d = 15$
0.29% ($\pm$0.05)	1.59% ($\pm$0.05)	4.36% ($\pm$2.61)	6.40% ($\pm$3.10)

Table 6.5: $\mathbb{C}$-uARIMA-IPA: Amari-index as a function of the dimension of the components d: average $\pm$ standard deviation. Sample size: $T = 30,000$. For other sample numbers, see Fig. 6.15(e).

and $L_e = 1$ we did not experience any degradation of performance up to $\lambda = 0.99$.

Chapter 7

Discussion

In this thesis we addressed the problem of searching for hidden independent multidimensional non-Gaussian components. Particularly, we

- introduced a general problem family called ARIMA-IPA (and PNL-ARMA-IPA), which cover extensions of the original ICA problem (post nonlinear-, autoregressive-, convolutive mixing, complex-valued variables), and
- derived separation principles (Chapter 4, Fig. 4.1) for this problem family lessening the i.i.d. assumption of the ISA Separation Theorem. Making use of these separation principles, the solution of the original problem can be reduced step-by-step to simpler subproblems.

Direct consequences of the separation principles:

- algorithmic *modularity*: the solution of the specific subproblems can be performed by *any* subproblem-solver algorithm. For example, in the 'uBSSD via LPA' technique (see Theorem 8) any AR fitting and ISA solver can be combined. Provided that we solve the obtained ISA problem by making use of the ISA Separation Theorem, any 'well-tried' ICA procedure can be applied, then the computation of the dependence of the estimated ICA elements can be carried out by a 'well-performing' mutual information estimating routine, and a 'favourite' clustering procedure can complete the solution by grouping the ICA elements.

- computationally efficient and robust methods can be constructed to the specific subproblems. Namely,
  - the CE technique can be tailored to the ISA subproblem (see Section 5.1) to get a reliable, global permutation search solution.
  - the JFD based ISA method (Section 5.2), which builds upon the joint decorrelation over a function set, make it possible to solve 'high dimensional' ISA problems: dimensionality can be $300 - 400$, which is a great improvement over any other methods that can only handle problems of 20 dimensions.
  - Non-combinatorial approximations can be derived for the estimation of the dimension of the hidden components in ISA-reducible models (Section 5.3).

Several questions remain open:

ISA Separation Theorem–Characterization, Extension: Our simulations indicate that the presently known sufficient conditions of the separation theorem may be extended considerably. For example, the following datasets used in the presented studies are all outside the problem domain: (i) Beatles- [9, 10], Led Zeppelin songs [8], (ii) self-similar structures [11], (iii) deterministic dynamical systems [8].

The validity domain, the characterization of the sources types satisfying the ISA Separation Theorem is an open question, only sufficient conditions are known (see Fig. 4.3 and Fig. 4.4). Similarly, it would be interesting to find sufficient conditions for the 'JBD Separation Principle' (see Note 3) formulated in terms of *components* and guaranteeing *global* extrema.

$\mathbb{C}$-ISA Separability: The $\mathbb{C}$-ISA problem can solved by $\mathbb{R}$-ISA provided that there is at least one Gaussian among the hidden components (see Section 4.4.1). However, the $\mathbb{C}$-ICA separability result of [68] suggests that the $\mathbb{C}$-ISA problem can also be solved for certain non-Gaussian hidden components. For $\mathbb{C}$-ISA we are not aware of any similar separability result. Thus, our $\mathbb{C}$-ISA Separation Theorem presented in Section 4.4.2 can be considered as a pioneering and preliminary result: its exact validity domain has not been characterized yet.

Amari-index: One can measure the efficiency of algorithms for solving ISA(-reducible) problems by the Amari-index. Our open questions:

- We conducted extended simulations to assess the efficiency, robustness and limits of our separation principles. According to our numerical experiments, the estimation error of our algorithms often (almost always) followed power law $r(T) \propto T^{-c}$ ($c > 0$). It would be interesting to find its optimality or characterize source types this relation holds for.
- It would be important to define a *normalized* performance measure similar to (6.1) for the general $d_m \geq 1$ case.

PNL-ISA *Component-Wise* Distortions: In the PNL-ISA problem, the $\mathbf{f}$ distortion were assumed to act *coordinate-wise*. Surely, this restriction can be relaxed to *components-wise* distortions, acting on d_m-dimensional coordinate sets.

Appendix A

Proofs

Proof I (ℂ-ISA Separability) *The ℂ-ISA problem can be reduced to ℝ-ISA, if there is at most one Gaussian component among $\varphi_v(\mathbf{e}^m)$, see Section 4.4.1. The ℝ-ISA separability result can be applied to the associated real problem. Finally, using that mapping $\varphi_M : Gl_{\mathbb{C}}(D_e) \to Gl_{\mathbb{R}}(2D_e)$ is a group homomorphism [see properties (A.22) and (A.23)] and injective ($Ker(\varphi_M) = \{\mathbf{I}\}$) $Gl_{\mathbb{C}}(D_x) \cong Im(\varphi_M)$ holds according to the group homomorphism theorem.*

Note II (ℝ-ISA: Undercomplete–Complete Reduction and Whitening via PCA) *In order to transform the undercomplete ℝ-ISA task into a complete ℝ-ISA task with white observations let $\mathbf{C} := cov[\mathbf{x}] = E[\mathbf{x}\mathbf{x}^T] = \mathbf{A}\mathbf{A}^T \in \mathbb{R}^{D_x \times D_x}$ denote the covariance matrix of the observation. Rank of $\mathbf{C}$ is D_e, since the rank of matrix $\mathbf{A}$ is D_e according to our assumptions. Matrix $\mathbf{C}$ is symmetric ($\mathbf{C} = \mathbf{C}^T$), thus it can be decomposed as follows: $\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, where $\mathbf{U} \in \mathbb{R}^{D_x \times D_e}$, and the columns of matrix $\mathbf{U}$ are orthogonal, that is, $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Furthermore, the rank of diagonal matrix $\mathbf{D} \in \mathbb{R}^{D_e \times D_e}$ is D_e. The principal component analysis can provide a decomposition in the desired form. Let $\mathbf{Q} := \mathbf{D}^{-1/2}\mathbf{U}^T \in \mathbb{R}^{D_e \times D_x}$. Then the original observation $\mathbf{x}$ can be modified to $\mathbf{x}' := \mathbf{Q}\mathbf{x} = \mathbf{Q}\mathbf{A}\mathbf{e} \in \mathbb{R}^{D_e}$. The resulting $\mathbf{x}'$ is* white *and can be regarded as the observation of a* complete *ISA task having mixing matrix $\mathbf{Q}\mathbf{A} \in \mathcal{O}^{D_e}$.*

Proof III (of Theorem 1 on page 19) *In the proof below, superscript is reserved for derivatives, and locally (instead of the former superscipts) subscripts are used for the notation of the m^{th} component.*

To prove the first statement it suffices to show that if $\mathbf{y} := \mathbf{Wh}(\mathbf{Ae})$ *is* $\mathbf{d}$*-independent then derivatives* h'_{mi} *are constant for* $\forall(i,m) \in \{1,\dots,d\} \times \{1,\dots,M\}$ *where* $\mathbf{h}_m$ *is part of* $\mathbf{h}$ *that belongs to subspace* m *and* h_{mi} *is the* i^{th} *coordinate function of* $\mathbf{h}_m : \mathbb{R}^d \to \mathbb{R}^d$*. It directly implies the second part of the statement as if* $\mathbf{Be}+\mathbf{Wp}$ *(where* $\mathbf{B}$ *denotes the matrix product* $\mathbf{WLA}$*) is* $\mathbf{d}$*-independent then* $\mathbf{Be}$ *is also* $\mathbf{d}$*-independent. In turn, because of the separability properties of the linear ISA,* $\mathbf{B}$ *can recover the hidden components up to permutation and invertible transformation within the subspaces (and maybe up to a constant translation within the subspaces).*

To prove the first part, let p_Y *and* p_E *denote the density functions of* $\mathbf{y}$ *and* $\mathbf{e}$*, respectively. Based on the transformation rule of the density functions,* p_Y *can be given as*

$$p_Y[\mathbf{Wh}(\mathbf{Ae})] = |\det(\mathbf{W})|^{-1} \left(\prod_{m=1}^{M} |\det(\mathbf{h}'_m[(\mathbf{Ae})_m])|^{-1} \right) |\det(\mathbf{A})|^{-1} p_E(\mathbf{e}), \quad \text{(A.1)}$$

where $(\mathbf{Ae})_m \in \mathbb{R}^d$ *is part of* $\mathbf{Ae}$ *belonging to subspace* m *(*$\mathbf{Ae} = [\mathbf{Ae}_1;\dots;\mathbf{Ae}_M] \in \mathbb{R}^{Md}$*). In addition,* $\mathbf{y}$ *is* $\mathbf{d}$*-independent implying that* $p_Y(\cdot)$ *is* d-separated*, that is it can be rewritten as*

$$p_Y = \otimes_{m=1}^{M} p_m, \quad \text{(A.2)}$$

where functions p_m *are of* $\mathbb{R}^d \to \mathbb{R}$*. Let us choose point* $\mathbf{e}^0$*, for which* $p_E(\mathbf{e}^0) > 0$*. Then, there exists an open neighborhood* $U \subseteq \mathbb{R}^{D_e}$ *of this point, where* $p_E|_U > 0$*, and* $p_E|_U \in C^2(U,\mathbb{R})$*. Let us define* $r(\mathbf{e}) := \ln\left[|\det(\mathbf{W})|^{-1}|\det(\mathbf{A})|^{-1}p_E(\mathbf{e})\right]$ *on set* U*. It can be shown using Eqs.* (A.1)-(A.2) *that the following relation holds*

$$r(\mathbf{e}) = \ln\left[\left(\prod_{m=1}^{M} |\det(\mathbf{h}'_m[(\mathbf{Ae})_m])|\right)\left(\prod_{m=1}^{M} p_m([\mathbf{Wh}(\mathbf{Ae})]_m)\right)\right] \quad \text{(A.3)}$$

$$= \sum_{m=1}^{M} \ln[|\det(\mathbf{h}'_m[(\mathbf{Ae})_m])|] + \xi_m([\mathbf{Wh}(\mathbf{Ae})]_m), \quad (\mathbf{e} \in U) \quad \text{(A.4)}$$

where $\xi_m := \ln(p_m)$*, locally at* $\mathbf{e}^0_m$*.* p_E *is* d*-separated, thus function* $r \in C^2(U,\mathbb{R})$ *is* linearly d-separated*. In other words, it can be written as a direct sum, so* $\partial_i\partial_j r \equiv 0$*, where* $\lfloor \frac{i}{d} \rfloor \neq \lfloor \frac{j}{d} \rfloor$ *(*i*,*j *correspond to indices in different subspaces). However, since* p_m *and therefore* $\xi_m = \ln(p_m)$ *is locally constant, this relation holds (without loss of gen-*

*erality on set U) when ignoring terms ξ_m, since their derivatives are 0. Now, following the reasoning of Lemma 3.5 [171] for the d-separated function $\left[\otimes_{m=1}^{M} \det(\mathbf{h}'_m)\right]$ (**As**), where $\mathbf{A}$ is 'mixing' ($|\cdot|$s were dropped since functions $\mathbf{h}_m$ were assumed to be invertible), we can see that each function $g_m(\mathbf{v}) = \det[\mathbf{h}'_m(\mathbf{v})] (\not\equiv 0)$ satisfies a differential equation*

$$g_m H_{g_m} - \nabla g_m (\nabla g_m)^T \equiv \mathbf{C}_m g_m^2 \tag{A.5}$$

on set $U_m := \mathbf{A}_m(U)$, where $\mathbf{A}_m := [\mathbf{A}_{m1}, \dots, \mathbf{A}_{mM}]$, $\mathbf{C}_m \in \mathbb{R}^{d\times d}$, ∇ stands for the gradient, H is the Hessian. For this reason—similarly to Lemma 3.4 [171]—functions g_m can be given in the form $g_m(\mathbf{v}) = e^{\mathbf{v}^T\mathbf{D}_m\mathbf{v}+\mathbf{b}_m^T\mathbf{v}+c_m}$ $(\mathbf{v} \in U_m)$ with suitable $\mathbf{D}_m \in \mathbb{R}^{d\times d}, \mathbf{b}_m \in \mathbb{R}^d, c_m \in \mathbb{R}$. Furthermore, as functions $\mathbf{h}_m$ are assumed to act on each coordinate separately as h_{mi}, g_m can be written as $g_m = \otimes_{i=1}^{d} h'_{mi}$, we arrive at

$$h'_{mi}(t) = \pm e^{d_{mi}t^2+b_{mi}t+c_{mi}} \quad (d_{mi}, b_{mi}, c_{mi} \in \mathbb{R}) \tag{A.6}$$

locally, exploiting that $h'_{mi} \equiv 0$ is not allowed. Based on our assumption on h_{mi} it holds on all $\mathbb{R}(\ni t)$.

Let us introduce the following notation: $\mathbf{z} = [\mathbf{z}_1; \dots; \mathbf{z}_M] = \mathbf{A}\mathbf{e}$, where $\mathbf{z} \in \mathbf{A}(U)$. Following the above reasoning for the inverse system $\mathbf{e} = \mathbf{A}^{-1}\mathbf{h}^{-1}(\mathbf{W}^{-1}\mathbf{z})$, h_{mi}^{-1} can be given similar to (A.6), with other constants. However, if both $h_{mi}{}'$ and $(h_{mi}^{-1})'$ are of exponential type then it follows that $d_{mi} = 0$ and $b_{mi} = 0$ [$\forall(m,i) \in \{1,\dots,M\} \times \{1,\dots,d\}$]. Therefore, h_{mi}s are affine, that is $h_{mi}(u) = l_{mi}(u) + p_{mi}$ ($l_{mi}, p_{mi} \in \mathbb{R}$), what we wanted to demonstrate.

Proof IV (of Theorem 5 on page 25) *Below we prove that the $\mathbb{R}$-w-EPI condition (4.7) [which implies (4.1)] is valid for spherical variables, and for 2-dimensional variables invariant to 90° rotation. In what follows, $\mathbf{u} = \mathbf{e}^m \in \mathbb{R}^{d_m}$ stands for the m^{th} hidden component.*

Spherical variables: *For spherical variables, the distribution and thus the entropy of these projections are independent of $\mathbf{w} \in \mathcal{S}_{\mathbb{R}}^{d_m}$. Because $e^{2H(w_i u_i)} = e^{2H(u_i)} w_i^2$ and $\mathbf{w} \in \mathcal{S}_{\mathbb{R}}^{d_m}$, the $\mathbb{R}$-w-EPI is satisfied with equality $\forall \mathbf{w} \in \mathcal{S}_{\mathbb{R}}^{d_m}$.*

2-dimensional variables invariant to 90° rotation: *Assume that function*

$$f : \mathcal{S}^2 \ni \mathbf{w} \mapsto H\left(\sum_{i=1}^{2} w_i u_i\right) \tag{A.7}$$

has global minimum on set $\mathcal{S}^2_{\mathbb{R}} \cap \{\mathbf{w} \geq \mathbf{0}\}$.[1] *Let this minimum be at $\mathbf{w}_m \in \mathbb{R}^2$. Then, the 90° invariance warrants that function f take its global minimum also on $\mathbf{w}_m^\perp \in \mathbb{R}^2$, which is perpendicular to $\mathbf{w}_m$. Let $(\mathbf{C}^m)^T = [\mathbf{w}_m, \mathbf{w}_m^\perp] \in \mathcal{O}^2$. Now, we can estimate variables $\mathbf{C}^m\mathbf{e}^m$. This is sufficient because the $\mathbb{R}$-ISA solution is ambiguous up to orthogonal transformation within the subspaces.*

Proof V (of Theorem 6 on page 27) *Let L' be such that*

$$D_x L' \geq D_e(L_e + L') \tag{A.8}$$

is fulfilled. Such L' exists due to the undercomplete assumption $D_x > D_e$:

$$L' \geq \left\lceil \frac{D_e L_e}{D_x - D_e} \right\rceil. \tag{A.9}$$

This choice of L' guarantees that the reduction gives rise to an (under)complete $\mathbb{R}$-ISA task: let $x_m(t)$ denote the m^{th} coordinate of observation $\mathbf{x}(t)$ and let the matrix $\mathbf{H}_l \in \mathbb{R}^{D_x \times D_e}$ be decomposed into $1 \times d_m$ sized blocks. That is, $\mathbf{H}_l = [\mathbf{H}_l^{ij}]_{i=1..D_x, j=1..M}$ ($\mathbf{H}_l^{ij} \in \mathbb{R}^{1 \times d_j}$), where i and j denote row and column indices, respectively. Using

[1] Relation $\mathbf{w} \geq \mathbf{0}$ concerns each coordinate.

notations

$$\mathbf{E}^m(t) := [\mathbf{e}^m(t); \mathbf{e}^m(t-1); \ldots; \mathbf{e}^m(t-(L_e+L')+1)] \in \mathbb{R}^{d_m(L_e+L')}, \tag{A.10}$$

$$\mathbf{X}^m(t) := [x_m(t); x_m(t-1); \ldots; x_m(t-L'+1)] \in \mathbb{R}^{L'}, \tag{A.11}$$

$$\mathbf{E}(t) := [\mathbf{E}^1(t); \ldots; \mathbf{E}^M(t)] \in \mathbb{R}^{\sum_{m=1}^{M} d_m(L_e+L')=D_e(L_e+L')}, \tag{A.12}$$

$$\mathbf{X}(t) := [\mathbf{X}^1(t); \ldots; \mathbf{X}^{D_x}(t)] \in \mathbb{R}^{D_x L'}, \tag{A.13}$$

$$\mathbf{A}^{ij} := \begin{bmatrix} \mathbf{H}_0^{ij} & \ldots & \mathbf{H}_L^{ij} & \mathbf{0} & \ldots & \mathbf{0} \\ & \ddots & & \ddots & & \\ & & \ddots & & \ddots & \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{H}_0^{ij} & \ldots & \mathbf{H}_L^{ij} \end{bmatrix} \in \mathbb{R}^{L' \times d_j(L_e+L')}, \tag{A.14}$$

$$\mathbf{A} := [\mathbf{A}^{ij}]_{i=1..D_x, j=1..M} \in \mathbb{R}^{D_x L' \times \sum_{m=1}^{M} d_m(L_e+L')=D_x L' \times D_e(L_e+L')}, \tag{A.15}$$

model

$$\mathbf{X}(t) = \mathbf{A}\mathbf{E}(t) \tag{A.16}$$

can be obtained. Here, $\mathbf{e}^m(t)$*s are i.i.d. in time* t*, they are independent for different* m *values, and Equation* (A.8) *holds for* L'*. Thus,* (A.16) *is either an undercomplete or a complete* $\mathbb{R}$*-ISA task, depending on the relation of the l.h.s and the r.h.s of* (A.8)*: the task is complete if the two sides are equal. In* (A.16)*, each* $\mathbf{e}^m \in \mathbb{R}^{d_m}$ *component appears* $L_e + L'$ *times.*

Proof VI (of Theorem 8 (and its special case, Theorem 7) on page 29 (page 28)) *We assumed that* $\mathbf{H}[z]$ *has left inverse, thus the hidden* $\mathbf{e}$ *can be expressed from observation* $\mathbf{x}$ *by causal FIR filtering, i.e.,*

$$\mathbf{e} = \mathbf{H}^{-1}[z]\mathbf{F}[z]\mathbf{A}^{-1}\mathbf{x} = \mathbf{H}'[z]\mathbf{x}, \tag{A.17}$$

where $\mathbf{H}^{-1}[z] = \sum_{n=0}^{N} \mathbf{M}_n z^{-n} \in \mathbb{R}[z]^{D_e \times D_s}$*,* N *denotes the degree of the* $\mathbf{H}^{-1}[z]$ *polynomial,* $\mathbf{A}^{-1}$ *is the left inverse of matrix* $\mathbf{A}$ *and* $\mathbf{H}'[z] = \mathbf{H}^{-1}[z]\mathbf{F}[z]\mathbf{A}^{-1}$*. According*

to Eqs. (3.3)-(3.4) *the following equality chain holds*

$$\mathbf{x} = \mathbf{A}(\mathbf{I} - \mathbf{F}[z])\mathbf{A}^{-1}\mathbf{x} + \mathbf{A}\mathbf{H}[z]\mathbf{e} \tag{A.18}$$

$$= \mathbf{A}(\mathbf{I} - \mathbf{F}[z])\mathbf{A}^{-1}\mathbf{x} + [\mathbf{A}\mathbf{H}_0 + \mathbf{A}(\mathbf{H}[z] - \mathbf{H}_0)]\mathbf{e} \tag{A.19}$$

$$= \left[\mathbf{A}(\mathbf{I} - \mathbf{F}[z])\mathbf{A}^{-1} + \mathbf{A}(\mathbf{H}[z] - \mathbf{H}_0)\mathbf{H}'[z]\right]\mathbf{x} + \mathbf{A}\mathbf{H}_0\mathbf{e}. \tag{A.20}$$

Here: in (A.19) *the term containing* $\mathbf{H}_0$ *was separated, then expression* (A.17) *for* $\mathbf{e}$ *was exploited. The first term of the observation* $\mathbf{x}$ *in* (A.20) *belongs to the linear hull of the finite history of* $\mathbf{x}$*, and using that the constant parts of* $\mathbf{I} - \mathbf{F}[z]$ *and* $\mathbf{H}[z] - \mathbf{H}_0$ *(containing* z^0*) cancelled out:*

$$\left[\mathbf{A}(\mathbf{I} - \mathbf{F}[z])\mathbf{A}^{-1} + \mathbf{A}(\mathbf{H}[z] - \mathbf{H}_0)\mathbf{H}'[z]\right]\mathbf{x} \in \left\langle z\mathbf{x}, z^2\mathbf{x}, \ldots, z^{\max(L_s, L_e+N)}\mathbf{x}\right\rangle. \tag{A.21}$$

Because $\mathbf{e}(t)$ *is independent of* $\langle \mathbf{x}(t-1), \mathbf{x}(t-2), \ldots, \mathbf{x}(t - \max(L_s, L_e + N))\rangle$*, we have that observation process* $\mathbf{x}(t)$ *is autoregressive with innovation* $\mathbf{A}\mathbf{H}_0\mathbf{e}(t)$*.*

Proof VII (see Section 4.2.2, page 29)

1. *Let us notice that instead of* $\mathbf{E}(t)$*,* $\mathbf{X}(t)$*, we could choose concatenations* $\mathbf{E}'(t) = [\mathbf{e}(t); \ldots; \mathbf{e}(t-(L_e+L')+1)]$*,* $\mathbf{X}'(t) = [\mathbf{x}(t); \ldots; \mathbf{x}(t-L'+1)]$*, too. Permutation is the only difference between these two quantities, thus one can arrive at a similar* $\mathbf{X}'(t) = \mathbf{A}'\mathbf{E}'(t)$ *model. In what follows, this permuted form will be denoted by* $\mathbf{X}(t)$*,* $\mathbf{E}(t)$*.*

2. *Hidden source* $\mathbf{e}$ *is estimated [in* $K := (L_e + L')$ *pieces] by the TCC method—, possibly after PCA computation for the undercomplete case—as* $\hat{\mathbf{E}} = [\hat{\mathbf{e}}_1^1; \ldots; \hat{\mathbf{e}}_K^1; \ldots; \hat{\mathbf{e}}_1^M; \ldots; \hat{\mathbf{e}}_K^M] = \hat{\mathbf{W}}^{\mathrm{TCC}}\mathbf{X}$*, where* $\hat{\mathbf{W}}^{\mathrm{TCC}} := \hat{\mathbf{W}}_{\mathrm{ISA}}^{\mathrm{TCC}}\hat{\mathbf{W}}_{\mathrm{PCA}}^{\mathrm{TCC}}$*. Now, one can take this expression at time* t*, use the definitions of* $\mathbf{E}$ *and* $\mathbf{X}$*, decompose demixing matrix* $\hat{\mathbf{W}}^{\mathrm{TCC}}$ *as* $\hat{\mathbf{W}}^{\mathrm{TCC}} = [\hat{\mathbf{W}}_{m,k}^{\mathrm{TCC},l}]_{l=0,\ldots,L'-1;m=1,\ldots,M;k=1,\ldots,K}$ *and have that* $\hat{\mathbf{e}}_k^m(t) = \sum_{l=0}^{L'-1}\hat{\mathbf{W}}_{m,k}^{\mathrm{TCC},l}\mathbf{x}(t-l)$ *with matrices* $\hat{\mathbf{W}}_{m,k}^{\mathrm{TCC},l} \in \mathbb{R}^{d_m \times D_x}$*. Introducing polynomial matrix* $\hat{\mathbf{W}}_{m,k}^{\mathrm{TCC}}[z] := \sum_{l=0}^{L'-1}\hat{\mathbf{W}}_{m,k}^{\mathrm{TCC},l}z^l$*, the above expression takes the following form:* $\hat{\mathbf{e}}_k^m = \hat{\mathbf{W}}_{m,k}^{\mathrm{TCC}}[z]\mathbf{x}$*.*

Proof VIII (of Theorem 9 on page 32) *Known properties of mappings* φ_v*,* φ_M

are as follows [114]:

$$\det[\varphi_M(\mathbf{M})] = |\det(\mathbf{M})|^2 \qquad (\mathbf{M} \in \mathbb{C}^{L\times L}), \tag{A.22}$$

$$\varphi_M(\mathbf{M}_1\mathbf{M}_2) = \varphi_M(\mathbf{M}_1)\varphi_M(\mathbf{M}_2) \qquad (\mathbf{M}_1 \in \mathbb{C}^{L_1\times L_2}, \mathbf{M}_2 \in \mathbb{C}^{L_2\times L_3}), \tag{A.23}$$

$$\varphi_v(\mathbf{M}\mathbf{v}) = \varphi_M(\mathbf{M})\varphi_v(\mathbf{v}) \qquad (\mathbf{M} \in \mathbb{C}^{L_1\times L_2}, \mathbf{v} \in \mathbb{C}^{L_2}), \tag{A.24}$$

$$\varphi_M(\mathbf{M}_1 + \mathbf{M}_2) = \varphi_M(\mathbf{M}_1) + \varphi_M(\mathbf{M}_2) \qquad (\mathbf{M}_1, \mathbf{M}_2 \in \mathbb{C}^{L_1\times L_2}), \tag{A.25}$$

$$\varphi_M(c\mathbf{M}) = c\varphi_M(\mathbf{M}) \qquad (\mathbf{M} \in \mathbb{C}^{L_1\times L_2}, c \in \mathbb{R}). \tag{A.26}$$

In words: (A.22) *describes transformation of determinant, while* (A.23), (A.24), (A.25) *and* (A.26) *expresses preservation of operation for matrix-matrix multiplication, matrix-vector multiplication, matrix addition, real scalar-matrix multiplication, respectively.*

Now, one may apply φ_v *to the* (3.1)-(3.2) $\mathbb{C}$*-uARIMA-IPA equations (with* $\mathbb{K} = \mathbb{C}$*) and use* (A.24)-(A.26)*. The result is as follows:*

$$\varphi_M(\mathbf{F}[z])\nabla^r[z]\varphi_v(\mathbf{s}) = \varphi_M(\mathbf{H}[z])\varphi_v(\mathbf{e}), \tag{A.27}$$

$$\varphi_v(\mathbf{x}) = \varphi_M(\mathbf{A})\varphi_v(\mathbf{s}). \tag{A.28}$$

Given that (i) the independence of $\mathbf{e}^m \in \mathbb{C}^{d_m}$ *is equivalent to that of* $\varphi_v(\mathbf{e}^m) \in \mathbb{R}^{2d_m}$*, and (ii) the stability of* $\varphi_M(\mathbf{F}[z])$ *and the existence of the left inverse of* $\varphi_M(\mathbf{H}[z])$ *are inherited from* $\mathbf{F}[z]$ *and* $\mathbf{H}[z]$*, respectively [see Eqs.* (A.22) *and* (A.23)*], we end up with an* $\mathbb{R}$*-uARIMA-IPA task with* (L_s, L_e, r) *parameters and* M *pieces of* $2d_m$*-dimensional hidden components* $\varphi_v(\mathbf{e}^m)$*.*

Proof IX (of Theorem 10 on page 32) *These are the main steps of the proof:*

1. *First, we show that the mutual information based* $\mathbb{C}$*-ISA cost function is equivalent to the minimization of multidimensional entropies, which is more appropriate to the rest of the proof.*

2. *Then, we define a so-called* $\mathbb{C}$*-EPI property and a* $\mathbb{C}$*-w-EPI relation starting from the vector variant of the* $\mathbb{R}$*-EPI relation.*

3. *The next step is to prove that the cost function* $\mathbb{C}$*-ICA—assuming the* $\mathbb{C}$*-w-EPI property for the hidden sources—has global minimum in the* $\mathbb{C}$*-ISA solution.*

4. Finally, the invariance of the $\mathbb{C}$-ICA cost is used.

In detail:

1. *The $\mathbb{C}$-ISA task can be viewed as the minimization of the mutual information between the estimated components on the unitary group [see (3.11)]:*

$$J_{\mathrm{I},\mathbb{C}\text{-ISA}}(\mathbf{W}) := I(\mathbf{y}^1, \ldots, \mathbf{y}^M) \to \min_{\mathbf{W} \in \mathcal{U}^{D_e}}, \tag{A.29}$$

where $\mathbf{y} = \mathbf{W}\mathbf{x}$, $\mathbf{y} = \left[\mathbf{y}^1; \ldots; \mathbf{y}^M\right]$, $\mathbf{y}^m \in \mathbb{R}^{d_m}$, $D_e = \sum_{m=1}^{M} d_m$. Mutual information (Shannon's multidimensional entropy) of complex variables is defined as the mutual information (entropy) of their images under mapping φ_v [68]:

$$I(\mathbf{y}^1, \ldots, \mathbf{y}^M) := I(\varphi_v(\mathbf{y}^1), \ldots, \varphi_v(\mathbf{y}^M)), \tag{A.30}$$

$$H(\mathbf{y}) := H(\varphi_v(\mathbf{y})). \tag{A.31}$$

Matrix $\mathbf{W}$ is unitary, thus $\varphi_M(\mathbf{W})$ is orthogonal [see (A.22)] and the $\mathbb{C}$-ISA task is equivalent to the minimization of the cost function

$$J_{\mathbb{C}\text{-ISA}}(\mathbf{W}) := \sum_{m=1}^{M} H\left(\mathbf{y}^m\right) \to \min_{\mathbf{W} \in \mathcal{U}^{D_e}}, \tag{A.32}$$

because:

$$I\left(\mathbf{y}^1,\ldots,\mathbf{y}^M\right) = I\left(\varphi_v(\mathbf{y}^1),\ldots,\varphi_v(\mathbf{y}^M)\right) \tag{A.33}$$

$$= \sum_{m=1}^{M} H\left(\varphi_v(\mathbf{y}^m)\right) - H(\varphi_v(\mathbf{y})) \tag{A.34}$$

$$= \sum_{m=1}^{M} H\left(\varphi_v(\mathbf{y}^m)\right) - H(\varphi_v(\mathbf{W}\mathbf{x})) \tag{A.35}$$

$$= \sum_{m=1}^{M} H\left(\varphi_v(\mathbf{y}^m)\right) - H(\varphi_M(\mathbf{W})\varphi_v(\mathbf{x})) \tag{A.36}$$

$$= \sum_{m=1}^{M} H\left(\varphi_v(\mathbf{y}^m)\right) - [H(\varphi_v(\mathbf{x})) + \ln(|\det(\varphi_M(\mathbf{W}))|)] \tag{A.37}$$

$$= \sum_{m=1}^{M} H\left(\varphi_v(\mathbf{y}^m)\right) - [H(\varphi_v(\mathbf{x})) + \ln(|\det(\mathbf{W})|^2)] \tag{A.38}$$

$$= \sum_{m=1}^{M} H\left(\varphi_v(\mathbf{y}^m)\right) - constant + 0 \tag{A.39}$$

$$= \sum_{m=1}^{M} H\left(\mathbf{y}^m\right) - constant. \tag{A.40}$$

Here:

(a) (A.33)*: definition of the complex mutual information was applied,*

(b) (A.34)*: we used a well-known relation between I and H [58],*

(c) (A.35)*: definition of $\mathbf{y}$ was plugged in,*

(d) (A.36)*: is the consequence of* (A.24)*,*

(e) (A.37)*: identity for the transformation of Shannon's differential entropy under linear mappings [58] is made use of,*

(f) (A.38)*: is the consequence of* (A.22)*,*

(g) (A.39)*: $H(\varphi_v(\mathbf{x}))$ is constant in $\mathbf{W}$, and $|det(\mathbf{W})| = 1$ because matrix $\mathbf{W}$ is orthogonal,*

(h) (A.40)*: definition of the complex entropy is used.*

2. *EPI-type Relations: Let us consider the vector variant of the* $\mathbb{R}$*-EPI relation.*

Lemma X (vector-EPI) *For independent (finite covariance random variables)* $\mathbf{u}_1, \ldots, \mathbf{u}_L \in \mathbb{R}^q$ *holds [166] that*

$$e^{2H\left(\sum_{i=1}^{L} \mathbf{u}_i\right)/q} \geq \sum_{i=1}^{L} e^{2H(\mathbf{u}_i)/q}. \tag{A.41}$$

Let us define a similar property for complex random variables:

Definition XI ($\mathbb{C}$-EPI) *We say that random variables* u_1,...,$u_L \in \mathbb{C}$ *satisfy relation* $\mathbb{C}$*-EPI if*

$$e^{H\left(\sum_{i=1}^{L} u_i\right)} \geq \sum_{i=1}^{L} e^{H(u_i)}. \tag{A.42}$$

Note XII ($\mathbb{C}$-EPI - independence is sufficient) *This holds for independent random variables* $u_1, \ldots, u_L \in \mathbb{C}$*, because according to vector-EPI (*$q = 2$*)*

$$e^{2H\left(\sum_{i=1}^{L} u_i\right)/2} \geq \sum_{i=1}^{L} e^{2H(u_i)/2}. \tag{A.43}$$

We need the following lemma:

Lemma XIII ($\mathbb{C}$-w-EPI $\Rightarrow$ (A.45)) *Let us assume that random variables* u_1, ..., $u_L \in \mathbb{C}$ *satisfy condition*

$$e^{H\left(\sum_{i=1}^{L} w_i u_i\right)} \geq \sum_{i=1}^{L} e^{H(w_i u_i)} \quad \forall \mathbf{w} = [w_1; \ldots; w_L] \in S_{\mathbb{C}}^L \tag{A.44}$$

that we shall call condition $\mathbb{C}$*-w-EPI. Then*

$$H\left(\sum_{i=1}^{L} w_i u_i\right) \geq \sum_{i=1}^{L} |w_i|^2 H(u_i) \quad \forall \mathbf{w} \in S_{\mathbb{C}}^L. \tag{A.45}$$

Proof XIV (of Lemma XIII) *Assume that* $\mathbf{w} \in S_{\mathbb{C}}^L$*. Applying* ln *on condition* (A.44)*, and using the monotonicity of the* ln *function, we can see that the*

first inequality is valid in the following inequality chain

$$H\left(\sum_{i=1}^{L} w_i u_i\right) \geq \ln\left(\sum_{i=1}^{L} e^{H(w_i u_i)}\right) = \ln\left(\sum_{i=1}^{L} e^{H(u_i)}|w_i|^2\right) \tag{A.46}$$

$$\geq \sum_{i=1}^{L} |w_i|^2 \ln\left(e^{H(u_i)}\right) = \sum_{i=1}^{L} |w_i|^2 H(u_i). \tag{A.47}$$

Then,

(a) we used the relation:

$$H(wu) = H(u) + \ln\left(|w|^2\right) \quad (w, u \in \mathbb{C}) \tag{A.48}$$

for the entropy of the transformed variable. Hence

$$e^{H(w_i u_i)} = e^{H(u_i)+\ln\left(|w_i|^2\right)} = e^{H(u_i)} e^{\ln\left(|w_i|^2\right)} = e^{H(u_i)}|w_i|^2. \tag{A.49}$$

(b) In the second inequality, we exploited the concavity of ln.

3. *The $\mathbb{C}$-ISA Separation Theorem will be a corollary of the following claim:*

Proposition XV *Let $\mathbf{y} = \left[\mathbf{y}^1; \ldots; \mathbf{y}^M\right] = \mathbf{y}(\mathbf{W}) = \mathbf{We}$, where $\mathbf{W} \in \mathcal{U}^{D_e}$, $\mathbf{y}^m$ is the estimation of the m^{th} component of the $\mathbb{C}$-ISA task. Let y_i^m be the i^{th} complex coordinate of the m^{th} component. Similarly, let e_i^m stand for the i^{th} coordinate of the m^{th} source. Let us assume that the $\mathbf{e}^m$ sources satisfy condition* (A.45)*. Then*

$$\sum_{m=1}^{M}\sum_{i=1}^{d_m} H\left(y_i^m\right) \geq \sum_{m=1}^{M}\sum_{i=1}^{d_m} H\left(e_i^m\right). \tag{A.50}$$

Proof XVI (of Proposition XV) *Let us denote the $(i,j)^{th}$ element of matrix $\mathbf{W}$ by $W_{i,j}$. Coordinates of $\mathbf{y}$ and $\mathbf{e}$ will be denoted by y_i and e_i, respectively. Let $\mathcal{G}^1, \ldots, \mathcal{G}^M$ denote the indices belonging to the $1^{st}, \ldots, M^{th}$. subspaces, that is, $\mathcal{G}^1 := \{1, \ldots, d_1\}, \mathcal{G}^2 := \{d_1+1, \ldots, d_1+d_2\}, \ldots, \mathcal{G}^M := \{D_e - d_M + 1, \ldots, D_e\}$. Now, writing the elements of the i^{th} row of matrix multiplication $\mathbf{y} = \mathbf{We}$, we*

have

$$y_i = \sum_{j \in \mathcal{G}^1} W_{i,j} e_j + \ldots + \sum_{j \in \mathcal{G}^M} W_{i,j} e_j \tag{A.51}$$

and thus,

$$H(y_i) = H\left(\sum_{m=1}^{M} \sum_{j \in \mathcal{G}^m} W_{i,j} e_j \right) \tag{A.52}$$

$$= H\left(\sum_{m=1}^{M} \left[\left(\sum_{l \in \mathcal{G}^m} |W_{i,l}|^2 \right)^{\frac{1}{2}} \frac{\sum_{j \in \mathcal{G}^m} W_{i,j} e_j}{\left(\sum_{l \in \mathcal{G}^m} |W_{i,l}|^2 \right)^{\frac{1}{2}}} \right] \right) \tag{A.53}$$

$$\geq \sum_{m=1}^{M} \left[\left(\sum_{l \in \mathcal{G}^m} |W_{i,l}|^2 \right) H\left(\frac{\sum_{j \in \mathcal{G}^m} W_{i,j} e_j}{\left(\sum_{l \in \mathcal{G}^m} |W_{i,l}|^2 \right)^{\frac{1}{2}}} \right) \right] \tag{A.54}$$

$$= \sum_{m=1}^{M} \left[\left(\sum_{l \in \mathcal{G}^m} |W_{i,l}|^2 \right) H\left(\sum_{j \in \mathcal{G}^m} \frac{W_{i,j}}{\left(\sum_{l \in \mathcal{G}^m} |W_{i,l}|^2 \right)^{\frac{1}{2}}} e_j \right) \right] \tag{A.55}$$

$$\geq \sum_{m=1}^{M} \left[\left(\sum_{l \in \mathcal{G}^m} |W_{i,l}|^2 \right) \sum_{j \in \mathcal{G}^m} \left| \frac{W_{i,j}}{\left(\sum_{l \in \mathcal{G}^m} |W_{i,l}|^2 \right)^{\frac{1}{2}}} \right|^2 H(e_j) \right] \tag{A.56}$$

$$= \sum_{m=1}^{M} \left[\sum_{j \in \mathcal{G}^m} |W_{i,j}|^2 H(e_j) \right] \tag{A.57}$$

The above steps can be justified as follows:

(a) (A.52)*: Eq.* (A.51) *was inserted into the argument of* H*.*

(b) (A.53)*: New terms were added for Lemma XIII.*

(c) (A.54)*: Sources* $\mathbf{e}^m$ *are independent of each other and this independence is preserved upon mixing* within *the subspaces, and we could also use Lemma XIII, because* $\mathbf{W}$ *is a unitary matrix.*

(d) (A.55)*: Nominators were transferred into the* $\sum_j$ *terms.*

(e) (A.56)*: Variables* $\mathbf{e}^m$ *satisfy condition* (A.45) *according to our assumptions.*

(f) (A.57)*: We simplified the expression after squaring.*

Using this inequality, summing it for i, exchanging the order of the sums, and making use of the unitary property of matrix $\mathbf{W}$, we have

$$\sum_{i=1}^{D_e} H(y_i) \geq \sum_{i=1}^{D_e} \left(\sum_{m=1}^{M} \left[\sum_{j \in \mathcal{G}^m} |W_{i,j}|^2 H\left(e_j\right) \right] \right) \tag{A.58}$$

$$= \sum_{m=1}^{M} \left[\sum_{j \in \mathcal{G}^m} \left(\sum_{i=1}^{D_e} |W_{i,j}|^2 \right) H\left(e_j\right) \right] \tag{A.59}$$

$$= \sum_{j=1}^{D_e} H(e_j). \tag{A.60}$$

4. *Having this proposition, now we prove the $\mathbb{C}$-ISA Separation Theorem.*

   Proof XVII (of Theorem 10 on page 32) *$\mathbb{C}$-ICA minimizes the l.h.s. of Eq. (A.50), that is, it minimizes $\sum_{m=1}^{M} \sum_{i=1}^{d_m} H\left(y_i^m\right)$. The set of minima is invariant to permutations and to multiplication of the coordinates by numbers with unit absolute value, and according to Proposition XV $\{e_i^m\}$ (i.e., the coordinates of the $\mathbb{C}$-ISA task) is among the minima.*

   We can disregard multiplications with unit absolute values, because unitary ambiguity within subspaces are present in the $\mathbb{C}$-ISA task.

Proof XVIII (of Theorem 11 on page 33)

Complex spherical variables: *For complex spherical variables, the distribution and thus the entropy of these projections are independent of $\mathbf{w} \in \mathcal{S}_{\mathbb{C}}^{d_m}$, using that $\bar{\mathbf{w}} \in S_{\mathbb{C}}^{d_m} \Leftrightarrow \mathbf{w} \in S_{\mathbb{C}}^{d_m}$ (conjugation preserves length). Because $e^{H(w_i u_i)} = e^{H(u_i)} |w_i|^2$ and $\mathbf{w} \in \mathcal{S}_{\mathbb{C}}^{d_m}$, the $\mathbb{C}$-w-EPI is satisfied with equality $\forall \mathbf{w} \in \mathcal{S}_{\mathbb{C}}^{d_m}$.*

$\mathbb{C}$-w-EPI: *see Lemma XIII.*

Proof XIX (of Theorem 12 on page 37) *The statement follows from the inequality related to the multidimensional Shannon differential entropy H: Let $\mathbf{v} = [\mathbf{v}^1; \ldots; \mathbf{v}^M] \in \mathbb{R}^{D_e}$ ($\mathbf{v}^m \in \mathbb{R}^{d_m}$) denote a random variable. Then*

$$H\left(\mathbf{v}^1, \ldots, \mathbf{v}^M\right) \leq \sum_{m=1}^{M} H\left(\mathbf{v}^m\right), \tag{A.61}$$

and equality holds iff $\mathbf{v}^m$*s are independent [58]. Hint: one can choose* $\mathbf{u}$ *as a normal random variable with covariance* $\mathbf{\Sigma}(\mathbf{f}, T)$ *and insert the expression of the entropy of normal variables. We have proven the following statement: Let* $\mathbf{\Sigma} \in \mathbb{R}^{D_e \times D_e}$ *be a positive semi-definite matrix, let* $\mathbf{\Sigma}^{m,m} \in \mathbb{R}^{d_m \times d_m}$ *denote the* m^{th} *block in the diagonal of matrix* $\mathbf{\Sigma}$*, and let* $D_e = \sum_{m=1}^{M} d_m$*. Then the function*

$$0 \leq J(\mathbf{\Sigma}) := -\frac{1}{2} \log \left[\frac{\det(\mathbf{\Sigma})}{\prod_{m=1}^{M} \det(\mathbf{\Sigma}^{m,m})} \right] \tag{A.62}$$

is 0 *iff* $\mathbf{\Sigma} = blockdiag(\mathbf{\Sigma}^{1,1}, \ldots, \mathbf{\Sigma}^{M,M})$.

Appendix B

Pseudocodes

Below, we provide the pseudocodes of the '$\mathbb{R}$-uMA-IPA via TCC' (Section 4.2.2), the '$\mathbb{R}$-uARMA-IPA via LPA' (Section 4.2.3) and the JFD (Section 5.2) methods.

Table B.1: $\mathbb{R}$-uMA-IPA via TCC: Pseudocode

Input of the algorithm
observation: $\{\mathbf{x}(t)\}_{t=1,\ldots,T}$
Optimization
Apply temporal concatenation of length L' [see (A.9)] to $\mathbf{x}$: $\mathbf{X}$
Reduce uISA to ISA (and whiten): $\mathbf{X}' = \hat{\mathbf{W}}_{\text{PCA}}\mathbf{X}$
Apply ISA to $\mathbf{X}'$: demixing matrix is $\hat{\mathbf{W}}_{\text{ISA}}$, estimated source is $\hat{\mathbf{E}} = \hat{\mathbf{W}}_{\text{ISA}}\mathbf{X}'$
Estimation
$\hat{\mathbf{W}}_{\mathbb{R}\text{-uMA-IPA}} = \hat{\mathbf{W}}_{\text{ISA}}\hat{\mathbf{W}}_{\text{PCA}}$
$\hat{\mathbf{E}} = [\hat{\mathbf{e}}_1^1; \ldots; \hat{\mathbf{e}}_K^1; \ldots; \hat{\mathbf{e}}_1^M; \ldots; \hat{\mathbf{e}}_K^M] = \hat{\mathbf{W}}_{\mathbb{R}\text{-uMA-IPA}}\mathbf{X} \quad (K = L_e + L')$

Table B.2: $\mathbb{R}$-uARMA-IPA via LPA: Pseudocode

Input of the algorithm
Observation: $\{\mathbf{x}(t)\}_{t=1,\ldots,T}$
Optimization
AR fit: for observation $\mathbf{x}$ estimate $\hat{\mathbf{W}}_{\mathrm{AR}}[z]$
Estimate innovation: $\tilde{\mathbf{x}} = \hat{\mathbf{W}}_{\mathrm{AR}}[z]\mathbf{x}$
Reduce uISA to ISA (and whiten): $\tilde{\mathbf{x}}' = \hat{\mathbf{W}}_{\mathrm{PCA}}\tilde{\mathbf{x}}$
Apply ISA to $\tilde{\mathbf{x}}'$: demixing matrix is $\hat{\mathbf{W}}_{\mathrm{ISA}}$
Estimation

$$\hat{\mathbf{W}}_{\mathbb{R}\text{-uARMA-IPA}}[z] = \hat{\mathbf{W}}_{\mathrm{ISA}}\hat{\mathbf{W}}_{\mathrm{PCA}}\hat{\mathbf{W}}_{\mathrm{AR}}[z]$$
$$\hat{\mathbf{e}} = \hat{\mathbf{W}}_{\mathbb{R}\text{-uARMA-IPA}}[z]\mathbf{x}$$

Table B.3: JFD Algorithm: Pseudocode

Input of the algorithm
observation: $\{\mathbf{x}(t)\}_{t=1,\ldots,T}$
Optimization[a]
Apply ICA to $\mathbf{x}$: ICA demixing matrix is $\hat{\mathbf{W}}_{\mathrm{ICA}}$,
estimated source is $\hat{\mathbf{e}}_{\mathrm{ICA}} = \hat{\mathbf{W}}_{\mathrm{ICA}}\mathbf{x}$
Permutation search
$\mathbf{P} := \mathbf{I}$
repeat
sequentially for $\forall p \in \mathcal{G}^{m_1}, q \in \mathcal{G}^{m_2}$ $(m_1 \neq m_2)$:
if $J_{\mathrm{JFD}}(\mathbf{P}_{pq}\mathbf{P}; \mathcal{F}, T) < J_{\mathrm{JFD}}(\mathbf{P}; \mathcal{F}, T)$
$\mathbf{P} := \mathbf{P}_{pq}\mathbf{P}$
end
until $J_{\mathrm{JFD}}(\cdot; \mathcal{F}, T)$ decreases in the *sweep* above
Estimation
$\hat{\mathbf{W}}_{\mathrm{ISA}} = \mathbf{P}\hat{\mathbf{W}}_{\mathrm{ICA}}$
$\hat{\mathbf{e}}_{\mathrm{ISA}} = \hat{\mathbf{W}}_{\mathrm{ISA}}\mathbf{x} = \mathbf{P}\hat{\mathbf{e}}_{\mathrm{ICA}}$

[a] Let $\mathcal{G}^1, \ldots, \mathcal{G}^M$ denote the indices of the $1^{st}, \ldots, M^{th}$ subspaces, i.e., $\mathcal{G}^m := \{\sum_{m=1}^{M-1} d_m + 1, \ldots, \sum_{m=1}^{M} d_m\}$ $(d_0 := 0)$, and permutation matrix $\mathbf{P}_{pq}$ exchanges coordinates p and q.

Appendix C

Abbreviations

A summary of the abbreviations appears in the table below.

Abbreviation	Meaning
APEX	Adaptive Principal component EXtractor
AR	AutoRegressive
ARIMA	Integrated ARMA
ARMA	AutoRegressive Moving Average
BSD	Blind Source Deconvolution
BSSD	Blind SubSpace Deconvolution
CE	Cross-Entropy
CSA	Colored Subspace Analysis
CVNN	Complex-Valued Neural Network
ECG	ElectroCardioGraphy
EASI	Equivariant Adaptive Separation via Independence
EEG	ElectroEncephaloGraphy
EPI	Entropy Power Inequality
fMRI	functional Magnetic Resonance Imaging
FIR	Finite Impulse Response
ICA	Independent Component Analysis
IFS	Iterated Function System
IFSA	Independent Feature Subspace Analysis

i.i.d	independent identically distributed
IPA	Independent Process Analysis
ISA	Independent Subspace Analysis
JADE	Joint Approximate Diagonalization of Eigen-matrices
JBD	Joint Block Diagonalization
JFD	Joint $\mathcal{F}$-Decorrelation
KC	Kernel Covariance
KGV	Kernel Generalized Variance
LPA	Linear Prediction Approximation
LQG	Linear Quadratic Gaussian
LES	Linear Equation System
MEG	MagnetoEncephaloGraphy
MA	Moving Average
MICA	Multidimensional Independent Component Analysis
ML	Maximum Likelihood
MPB	Minimal Polynomial Basis
MSOBI	Multidimensional SOBI
NGCA	Non-Gaussian Component Analysis
PCA	Principal Component Analysis
PNL	Post NonLinear
PoT	Product of Experts
QAM	Quadrature Amplitude Modulated
RADICAL	Robust, Accurate, Direct ICA aLgorithm
RNN	Recurrent Neural Network
SOBI	Second-Order Blind Identification
SUT	Strong-Uncorrelating Transform
TCC	Temporal ConCatenation
TICA	Topographic ICA
TSP	Travelling Salesman Problem
VDCA	Variance-Dependent Component Analysis
WSS	Wide-Sense Stationary

Short Summary (in English)

1. In my thesis, I formulated a unified framework for the search of independent components covering the classical assumptions of the literature (ICA, ISA, AR-IPA, BSD, PNL-ICA). For the derived problem family, I proved separation principles: using these principles the solution of the original problem can be decomposed and reduced to simple(r) subproblems. Namely,

   (a) I proved sufficient conditions for the ISA Separation Theorem (which makes it possible to solve the ISA problem by clustering ICA elements): (i) spherical symmetry, or (ii) in the case of 2-dimensional hidden components invariance to 90° rotation suffices.

   (b) I formulated a common extension of the ISA and BSD task, the MA-IPA problem (Moving Average Independent Process Analysis). I justified that in the undercomplete case the MA-IPA problem can be solved by ISA after (i) temporal concatenation, or (ii) linear prediction of the observation.

   (c) I unified and extended the MA-IPA and AR-IPA problems to non-stationary direction, this is the ARIMA-IPA (Integrated AutoRegressive Moving Average Independent Process Analysis) problem. I proved, that in the undercomplete case (uARIMA-IPA) its solution can be reduced to ISA by applying temporal differentiating and AR identification.

   (d) I proved the complex ISA Separation Theorem, and justified that it holds, e.g., for complex spherical variables. I showed an alternative solution by real techniques for the complex uARIMA-IPA problem family, provided that there is at least one Gaussian among the associated components.

   (e) I defined and provided a solution (reduction) method for the common extension of the PNL-ICA and ISA tasks, the PNL-ISA problem.

2. Making use of the former separation principles I derived efficient solution/approximation procedures. Namely,

(a) For finding the optimal permutation of the ICA elements to solve ISA (see the ISA Separation Theorem), I adapted the 'global' Cross Entropy optimization technique.

(b) I constructed the JFD (Joint $\mathcal{F}$-Decorrelation) ISA method, which can cope with ISA problems in dimension a magnitude larger, than former state-of-the-art ISA techniques.

(c) I showed a non-combinatorial approximation scheme for ISA-reducible problems, not requiring the a priori knowledge of the component dimensions.

Short Summary (in Hungarian)

1. Disszertációmban a független komponens keresés klasszikus irányait (ICA, ISA, AR-IPA, BSD, PNL-ICA) egységes keretbe foglaltam. A megfogalmazott feladatokra dekompozíciós elveket származtattam, amikkel megoldásuk egyszerűsíthető, már ismert feladatokra vezethető:

   (a) Beláttam, hogy az ISA szeparációs tétel – mely segítségével az ISA feladat ICA-ra, és az ICA elemek csoportosítására vezethető – teljesüléséhez elégséges: (i) szférikus szimmetria, sőt (ii) 2-dimenziós rejtett komponensek esetén 90° fokos elforgatásra való invariancia is.

   (b) Az ISA és BSD feladatok közös kiterjesztéseként definiáltam az MA-IPA (Moving Average Independent Process Analysis) feladatot. Igazoltam, hogy undercomplete esetben ez: (i) időbeli konkatenáció, avagy (ii) lineáris predikció segítségével megoldható, és ezt a két megközelítést közös FIR szűrős megfogalmazásba illesztettem.

   (c) Az MA-IPA és AR-IPA feladatokat egységesítettem, és kiterjesztettem nem-stacionárius irányban, ez az ARIMA-IPA probléma. Igazoltam, hogy a kapott általánosítás undercomplete esetben (uARIMA-IPA) időbeli differenciálás és AR identifikáció segítségével ISA-ra redukálható.

   (d) A komplex változós ISA feladatra beláttam egy komplex ISA szeparációs tételt, és igazoltam, hogy komplex szférikusan szimmetrikus változók eleget tesznek feltételrendszerének. Emellett bebizonyítottam, hogy komplex uARIMA-IPA problémacsalád valós technikákkal is megoldható – amennyiben az asszociált komponensek közt legfeljebb egy Gauss változó van.

   (e) Definiáltam és megoldási módszert, visszavezetési elvet adtam a PNL-ICA és ISA feladatok közös kiterjesztésére a PNL-ISA problémára.

2. Ezen visszavezetési elvekre építve hatékony közelítési eljárásokat származtattam. Nevezetesen:

(a) Az ISA szeparációs tétel optimális permutációjának kereséséhez a CE (cross entropy) „globális” optimalizálási módszert adaptáltam.

(b) Egy együttes nem-lineáris dekorrelációra (Joint $\mathcal{F}$-Decorrelation, JFD) építő, korábbi irodalombeli módszerekkel összevetve egy nagyságrenddel nagyobb feladatok megoldására képes ISA módszert adtam.

(c) A rejtett források ismeretét nem-igénylő (nem-kombinatorikus) közelítést mutattam a fenti, ISA-ra redukálható feladatokra.

Own References

[1] András Lőrincz and Zoltán Szabó. Neurally plausible, non-combinatorial iterative independent process analysis. *Neurocomputing - Letters*, 70(7-9):1569–1573, 2007.

[2] Barnabás Póczos, Zoltán Szabó, Melinda Kiszlinger, and András Lőrincz. Independent process analysis without a priori dimensional information. In Mike E. Davies, Christopher J. James, Samer A. Abdallah, and Mark D. Plumbley, editors, *Independent Component Analysis and Blind Signal Separation (ICA 2007)*, volume 4666 of *Lecture Notes in Computer Science*, pages 252–259, Berlin Heidelberg, 9-12 September 2007. Springer-Verlag.

[3] Zoltán Szabó and András Lőrincz. Real and complex independent subspace analysis by generalized variance. In *ICA Research Network International Workshop (ICARN 2006)*, pages 85–88, 18-19 September 2006.

[4] Zoltán Szabó and András Lőrincz. Complex independent process analysis. *Acta Cybernetica*, 2007. (submitted).

[5] Zoltán Szabó and András Lőrincz. Independent subspace analysis can cope with the „curse of dimensionality". *Acta Cybernetica (and presented at the Symposium of Intelligent Systems 2006)*, 18:213–221, 2007.

[6] Zoltán Szabó, Barnabás Póczos, and András Lőrincz. Cross-entropy optimization for independent process analysis. In Justinian Rosca, Deniz Erdogmus, José C. Príncipe, and Simon Haykin, editors, *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *Lecture Notes in Computer Science*, pages 909–916. Springer, 5-8 March 2006.

[7] Zoltán Szabó, Barnabás Póczos, and András Lőrincz. Separation theorem for $\mathbb{K}$-independent subspace analysis with sufficient conditions. Technical report, Eötvös Loránd University, Budapest, 2006. http://arxiv.org/abs/math.ST/0608100.

[8] Zoltán Szabó, Barnabás Póczos, and András Lőrincz. Auto-regressive independent process analysis without combinatorial efforts. *Pattern Analysis and Applications*, 2007. (accepted).

[9] Zoltán Szabó, Barnabás Póczos, and András Lőrincz. Undercomplete blind subspace deconvolution. *Journal of Machine Learning Research*, 8:1063–1095, 2007.

[10] Zoltán Szabó, Barnabás Póczos, and András Lőrincz. Undercomplete blind subspace deconvolution via linear prediction. In J.N. Kok et al., editor, *European Conference on Machine Learning (ECML 2007)*, volume 4701 of *Lecture Notes in Artificial Intelligence*, pages 740–747, Berlin Heidelberg, 17-21 September 2007. Springer-Verlag.

[11] Zoltán Szabó, Barnabás Póczos, Gábor Szirtes, and András Lőrincz. Post nonlinear independent subspace analysis. In J. Marques de Sá et al., editor, *International Conference on Artificial Neural Networks (ICANN 2007)*, volume 4668 of *Lecture Notes in Computer Science - Part I*, pages 677–686, Berlin Heidelberg, 9-13 September 2007. Springer-Verlag.

External References

[12] Karim Abed-Meraim and Adel Belouchrani. Algorithms for joint block diagonalization. In *European Signal Processing Conference (EUSIPCO 2004)*, pages 209–212, 2004.

[13] Karim Abed-Meraim, Philippe Loubaton, and Eric Moulines. A subspace algorithm for certain blind identification problems. *IEEE Transactions on Information Theory*, 43(2):693–696, March 1997.

[14] Sophie Achard and Christian Jutten. Identifiability of post nonlinear mixtures. *IEEE Signal Processing Letters*, 12(5):423– 426, May 2005.

[15] Shotaro Akaho, Yasuhiko Kiuchi, and Shinji Umeyama. MICA: Multimodal independent component analysis. In *International Joint Conference on Neural Networks (IJCNN '99)*, volume 2, pages 927–932, 1999.

[16] Ian F. Akyildiz, WellJan Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393–422, March 2002.

[17] Shun-ichi Amari, Andrzej Cichocki, and Howard H. Yang. Recurrent neural networks for blind separation of sources. In *International Symposium on Nonlinear Theory and its Applications NOLTA (NOLTA'95)*, pages 37–42, Bruges, Belgium, 18-21 October 1995.

[18] Shun-ichi Amari, Andrzej Cichocki, and Howard H. Yang. A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, 8:757–763, 1996.

[19] Jörn Anemüller. Second-order separation of multidimensional sources with constrained mixing system. In Justinian Rosca, Deniz Erdogmus, José C. Príncipe, and Simon Haykin, editors, *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *LNCS*, pages 16–23. Springer, 5-8 March 2006.

[20] Jörn Anemüller, Jeng-Ren Duann, Terrence J. Sejnowski, and Scott Makeig. Spatio-temporal dynamics in fMRI recordings revealed with complex independent component analysis. *Neurocomputing*, 69(13-15):1502–1512, 2006.

[21] Jörn Anemüller, Terrence J. Sejnowski, and Scott Makeig. Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Networks*, 16:1311–1323, 2003.

[22] Shoko Araki, Shoji Makino, Ryo Mukai, Tsuyoki Nishikawa, and Hiroshi Saruwatari. Fundamental limitation of frequency domain blind source separation for convolved mixture of speech. *IEEE Transactions on Speech and Audio Processing*, 11(2):109–116, March 2003.

[23] Francis R. Bach and Michael I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.

[24] Francis R. Bach and Michael I. Jordan. Beyond independent components: Trees and clusters. *Journal of Machine Learning Research*, 4:1205–1233, 2003.

[25] Andrew D. Back and Ah Chung Tsoi. Blind deconvolution of signals using a complex recurrent network. In *Neural Networks for Signal Processing*, pages 565–574. IEEE Press, 1994.

[26] Marian Stewart Bartlett, Javier R. Movellan, and Terrence J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, 2002.

[27] Michael Basin and Dario Calderon-Alvarez. Optimal LQG controller for linear stochastic systems with unknown parameters. *Journal of the Franklin Institute*, 345:293–302, 2008.

[28] Michael V. Basin and Irma R. Valadez Guzman. Optimal control in unobservable integral Volterra systems. *Journal of the Franklin Institute*, 339(1):13–27, 2002.

[29] Anthony J. Bell and Terrence J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37:3327–3338, 1997.

[30] Adel Belouchrani, Karim Abed-Meraim, Jean-François Cardoso, and Eric Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45:434–444, 1997.

[31] Alain Bensoussan. On the separation principle for distributed parameter systems. In *IFAC Conference on Control for Distributed Parameter Systems*, Banff, Canada, 1971.

[32] Alain Bensoussan and M. Viot. Optimal control of stochastic linear distributed parameter systems. *SIAM Journal on Control and Optimization*, 13:904–926, 1975.

[33] Ella Bingham and Aapo Hyvärinen. A fast fixed-point algorithm for independent component analysis of complex-valued signals. *International Journal of Neural Systems*, 10(1):1–8, 2000.

[34] Gilles Blanchard, Motoaki Kawanabe, Masashi Sugiyama, Vladimir Spokoiny, and Klaus-Robert Müller. In search of non-Gaussian components of a high-dimensional distribution. *Journal of Machine Learning Research*, 7:247–282, February 2006.

[35] Franco Blanchini, Stefano Miani, and Fouad Mesquine. A separation principle for linear switching systems and parametrization of all stabilizing controllers. 2008. (submitted), `http://users.dimi.uniud.it/~franco.blanchini/Blanchini_switch.pdf`.

[36] Hamid Bounit and H. Hammouri. A separation principle for distributed dissipative bilinear systems. *IEEE Transactions on Automatic Control*, 48(3):479–483, 2003.

[37] Luc Bouten and Ramon van Handel. On the separation principle of quantum control. Technical report, 2006. http://arxiv.org/abs/math-ph/0511021.

[38] Luc Bouten, Ramon van Handel, and Matthew James. An introduction to quantum filtering. *SIAM Journal on Control and Optimization*, 46:2199–2241, 2007.

[39] Vince D. Calhoun and Tülay Adali. Complex infomax: Convergence and approximation of infomax with complex nonlinearities. In *Neural Networks for Signal Processing*, 2002.

[40] Vince D. Calhoun and Tülay Adali. Complex infomax: Convergence and approximation of infomax with complex nonlinearities. *VLSI Signal Processing Systems for Signal, Image, and Video Technology*, 44(1/2):173–190, 2006.

[41] Jean-François Cardoso. Multidimensional independent component analysis. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, volume 4, pages 1941–1944, 1998.

[42] Jean-François Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.

[43] Jean-François Cardoso and Tülay Adali. The maximum likelihood approach to complex ICA. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006)*, volume 5, Toulouse, France, 14-19 March 2006.

[44] Jean-François Cardoso and Beate Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 1996.

[45] Jean-François Cardoso and Maude Martin. A flexible component model for precision ICA. In Mike E. Davies, Christopher J. James, Samer A. Abdallah, and Mark D. Plumbley, editors, *Independent Component Analysis and Signal Separation (ICA 2007)*, volume 4666 of *LNCS*, pages 1–8, Heidelberg, 2007. Springer.

[46] Jean-François Cardoso and Antoine Souloumiac. Blind beamforming for non-gaussian signals. *IEE-Proceedings-F*, 140(6):362–370, 1993.

[47] Marc Castella and Pierre Comon. Blind separation of instantaneous mixtures of dependent sources. In Mike E. Davies, Christopher J. James, Samer A. Abdallah, and Mark D. Plumbley, editors, *Independent Component Analysis and Signal Separation (ICA 2007)*, volume 4666 of *LNCS*, pages 9–16, Heidelberg, 2007. Springer.

[48] Yiu-ming Cheung and Lei Xu. Dual multivariate auto-regressive modeling in state space for temporal signal separation. *IEEE Transaction on Systems, Man, Cybernetics—Part B*, 33:386–398, 2003.

[49] Silvia Chiappa and David Barber. Generative temporal ICA for classification in asynchronous BCI systems. In *IEEE EMBS Conference On Neural Engineering*, pages 514–517, Arlington, VA, 16-19 March 2005.

[50] Heeyoul Choi and Seungjin Choi. Relative gradient learning for independent subspace analysis. In *International Joint Conference on Neural Networks (IJCNN 2006)*, pages 3919– 3924, Vancouver, Canada, 16-21 July 2006.

[51] Seungjin Choi. Differential learning algorithms for decorrelation and independent component analysis. *Neural Networks*, 19(10):1558–1567, December 2006.

[52] Seungjin Choi and Andrzej Cichocki. Blind signal deconvolution by spatio-temporal decorrelation and demixing. *Neural Networks for Signal Processing*, 7:426–435, 1997.

[53] Seungjin Choi, Andrzej Cichocki, Hyung-Min Park, and Soo-Yound Lee. Blind source separation and independent component analysis. *Neural Information Processing - Letters and Reviews*, 6:1–57, 2005.

[54] Andrzej Cichocki and Shun-ichi Amari. *Adaptive blind signal and image processing*. John Wiley & Sons, 2002.

[55] Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

[56] Oswaldo L.V. Costa and Marcelo D. Fragoso. A separation principle for the H_2-control of continuous-time infinite Markov jump linear systems with partial observations. *Journal on Mathematical Analysis and Applications*, 331:97–120, 2007.

[57] Oswaldo L.V. Costa and Esteban F. Tuesta. H_2-control and the separation principle for discrete-time Markovian jump linear systems. *Mathematics of Control, Signals and Systems*, 16:320–350, 2004.

[58] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. John Wiley and Sons, New York, USA, 1991.

[59] Sergio Cruces, Auxiliadora Sarmiento, and Iván Durán. The complex version of the minimum support criterion. In Mike E. Davies, Christopher J. James, Samer A. Abdallah, and Mark D. Plumbley, editors, *Independent Component Analysis and Signal Separation (ICA 2007)*, volume 4666 of *LNCS*, pages 17–24. Springer, Heidelberg, 2007.

[60] Nathalie Delfosse and Philippe Loubaton. Adaptive blind separation of convolutive mixtures. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP '96)*, pages 2940–2943, 1996.

[61] Fani Deligianni, Benny Lo, and Guang-Zhong Yang. Source recovery for body sensor network. In *International Workshop on Wearable and Implantable Body Sensor Networks 2006 (BSN 2006)*, pages 199–202, April 2006.

[62] Scott C. Douglas. Fixed-point algorithms for the blind separation of arbitrary complex-valued non-gaussian signal mixtures. *EURASIP Journal on Applied Signal Processing*, 2007(1), January 2007. Article ID 36525, 15 pages, doi:10.1155/2007/36525.

[63] Scott C. Douglas, Jan Eriksson, and Visa Koivunen. Equivariant algorithms for estimating the strong-uncorrelating transform in complex independent component analysis. In Justinian Rosca, Deniz Erdogmus, José C. Príncipe, and Simon Haykin, editors, *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *Lecture Notes in Computer Science*, pages 57–65, 5-8 March 2006.

[64] Scott C. Douglas, Jan Eriksson, and Visa Koivunen. Fixed-point complex ICA algorithms for the blind separation of sources using their real or imaginary components. In Justinian Rosca, Deniz Erdogmus, José C. Príncipe, and Simon Haykin, editors, *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *Lecture Notes in Computer Science*, pages 343–351, 5-8 March 2006.

[65] Scott C. Douglas, Hiroshi Sawada, and Shoji Makino. Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters. *IEEE Transactions on Speech and Audio Processing*, 13(1):92–104, January 2005.

[66] Mads Dyrholm, Scott Makeig, and Lars Kai Hansen. Model selection for convolutive ICA with an application to spatio-temporal analysis of EEG. *Neural Computation*, April 2007.

[67] Alan Edelman, Tomas Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[68] Jan Eriksson. Complex random vectors and ICA models: Identifiability, uniqueness and separability. *IEEE Transactions on Information Theory*, 52(3), 2006.

[69] Jan Eriksson and Visa Koivunen. Complex-valued ICA using second order statistics. In *Proceedings of the 14th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, pages 183– 192, 2004.

[70] Kai-Tai Fang, Samuel Kotz, and Kai Wang Ng. *Symmetric multivariate and related distributions*. Chapman and Hall, 1990.

[71] Cédric Févotte and Christian Doncarli. A unified presentation of blind source separation for convolutive mixtures using block-diagonalization. In *Independent Component Analysis and Blind Signal Separation (ICA 2003)*, pages 349–354, 2003.

[72] Simone Fiori. Blind separation of circularly distributed sources by neural extended APEX algorithm. *Neurocomputing Letters*, 34:239–252, 2000.

[73] Simone Fiori. Complex-weighted one-unit 'rigid-bodies' learning rule for independent component analysis. *Neural Processing Letters*, 15(3):275–282, 2002.

[74] Gary H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9:416–429, 1999.

[75] Alexei Gorokhov and Philippe Loubaton. Multiple-input multiple-output ARMA systems: Second order blind identification for signal extractions. In *IEEE Signal Processing Workshop on Statistical Signal and Array Processing (SSAP '96)*, pages 348 – 351, Washington, DC, USA, June 1996. IEEE Computer Society.

[76] Alexei Gorokhov and Philippe Loubaton. Subspace-based techniques for blind separation of convolutive mixtures with temporally correlated sources. *IEEE Transactions on Circuits and Systems–I Fundamental Theory and Applications*, 44(9):813 – 820, September 1997.

[77] Alexei Gorokhov and Philippe Loubaton. Blind identification of MIMO-FIR systems: A generalized linear prediction approach. *Signal Processing*, 73:105–124, 1999.

[78] Arthur Gretton, Alexander Smola, Olivier Bousquet, and Bernhard Schölkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6:2075–2129, December 2005.

[79] Michael J. Grimble. Stochastic control of discrete systems: A separation principle for Wiener and polynomial systems. *IEEE Transactions on Automatic Control*, 44(11):2125–2130, 1999.

[80] Duan Guang-Ren, S. Thompson, and Liu Guo-Ping. Separation principle for robust pole assignment – an advantage of full-order state observers. In *Proceedings of the 38th IEEE Conference on Decision and Control*, volume 1, pages 76–78, 1999.

[81] T.L. Gunckel and G.F. Francklin. A general solution for linear sampled data control. *Journal of Basic Engineering 85-D*, pages 197–201, 1963.

[82] Godfrey H. Hardy and Srinivasa I. Ramanujan. Asymptotic formulae in combinatory analysis. *Proceedings of the London Mathematical Society*, 17(1):75–115, 1918.

[83] John B. Hedgepeth, Vincent F. Gallucci, F. O'Sullivan, and Richard E. Thorne. An expectation maximization and smoothing approach for indirect acoustic estimation of fish size and density. *ICES Journal of Marine Science*, 56(1):36–50, 1999.

[84] Kenneth E. Hild, Hagai T. Attias, and Srikantan S. Nagarajan. An EM method for spatio-temporal blind source separation using an AR-MOG source model. In Justinian Rosca, Deniz Erdogmus, José C. Príncipe, and Simon Haykin, editors, *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *LNCS*, pages 98–105. Springer, 5-8 March 2006.

[85] Akira Hirose. *Complex-Valued Neural Networks: Theories and Applications*, volume 5 of *Series on Innovative Intelligence*. World Scientific Publishing Co. Pte. Ltd., 2004.

[86] Aapo Hyvärinen. Independent component analysis for time-dependent stochastic processes. In *International Conference on Artificial Neural Networks (ICANN '98)*, pages 541–546, Berlin, 1998. Springer-Verlag.

[87] Aapo Hyvärinen. Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11:1739–1768, 1999.

[88] Aapo Hyvärinen and Patrik O. Hoyer. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 12:1705–1720, 2000.

[89] Aapo Hyvärinen and Jarmo Hurri. Blind separation of sources that have spatiotemporal variance dependencies. *Signal Processing*, 84:247–254, 2004.

[90] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[91] Aapo Hyvärinen and Urs Köster. FastISA: A fast fixed-point algorithm for independent subspace analysis. In *European Symposium on Artificial Neural Networks (ESANN 2006)*, pages 371–376, Evere, Belgium, 2006. d-side.

[92] Aapo Hyvärinen and Erkki Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[93] Aapo Hyvärinen and Shohei Shimizu. A quasi-stochastic gradient algorithm for variance-dependent component analysis. In *International Conference on Artificial Neural Networks (ICANN 2006)*, pages 211–220, Athens, Greece, 2006.

[94] S. Icart and R. Gautier. Blind separation of convolutive mixtures using second and fourth order moments. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP '96)*, volume 5, pages 3018–3021, 1996.

[95] Alexander Ilin. Independent dynamics subspace analysis. In *European Symposium on Artificial Neural Networks (ESANN 2006)*, pages 345–350, April 2006.

[96] P.D. Joseph and J.T. Tou. On linear control theory. *AIEE Transactions, Applications and Industry*, 30:193–196, 1961.

[97] Tzyy-Ping Jung, Scott Makeig, Te-Won Lee, Martin J. McKeown, Glen Brown, Anthony J. Bell, and Terrence J. Sejnowski. Independent component analysis of biomedical signals. In *International Workshop on Independent Component Analysis and Signal Separation (ICA 2000)*, pages 633–644, June 2000.

[98] Christian Jutten and Jeanny Hérault. Blind separation of sources: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[99] Christian Jutten and Juha Karhunen. Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear systems. *International Journal of Neural Systems*, 14(5):267–292, 2004.

[100] R.E. Kalman and R.W. Koepecke. Optimal synthesis of linear sampling control systems using generalized performance indexes. *Transactions of the ASME*, pages 1820–1826, 1958.

[101] Yan Karklin and Michael S. Lewicki. A hierarchical Bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Computation*, 17(2):397 – 423, 2005.

[102] Hakan Karslı. Further improvement of temporal resolution of seismic data by autoregressive (AR) spectral extrapolation. *Journal of Applied Geophysics*, 59:324–336, 2006.

[103] Motoaki Kawanabe and Klaus-Robert Müller. Estimating functions for blind separation when sources have variance dependencies. *Journal of Machine Learning Research*, 6:453–482, 2005.

[104] Motoaki Kawanabe, Masashi Sugiyama, Gilles Blanchard, and Klaus-Robert Müller. A new algorithm of non-Gaussian component analysis with radial kernel functions. *Annals of the Institute of Statistical Mathematics*, 59(1):57–75, March 2007.

[105] Motoaki Kawanabe and Fabian J. Theis. Estimating non-Gaussian subspaces by characteristic functions. In Justinian Rosca, Deniz Erdogmus, José C. Príncipe, and Simon Haykin, editors, *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *Lecture Notes in Computer Science*, pages 157–164. Springer, 5-8 March 2006.

[106] Motoaki Kawanabe and Fabian J. Theis. Joint low-rank approximation for extracting non-Gaussian subspaces. *Signal Processing*, 87(8):1890–1903, August 2007.

[107] Heyjin Kim and Seungjin Choi. Independent subspaces of gene expression data. In *IASTED International Conference on Artificial Intelligence and Applications*, Innsbruck, Austria, 14-16 February 2005.

[108] Heyjin Kim, Seungjin Choi, and Sung-Yang Bang. Membership scoring via independent feature subspace analysis for grouping co-expressed genes. In *Inter-*

national Joint Conference on Neural Networks (IJCNN 2003), volume 3, pages 1690 – 1695, 20-24 July 2003.

[109] Jong Kyoung Kim and Seungjin Choi. Tree-dependent components of gene expression data for clustering. In *International Conference on Artificial Neural Networks (ICANN 2006)*, volume 4132 of *Lecture Notes in Computer Science*, pages 837–846. Springer Berlin / Heidelberg, 2006.

[110] Kimmo Kiviluoto and Erkki Oja. Independent component analysis for parallel financial time series. In *International Conference on Neural Information Processing (ICONIP '98)*, volume 2, pages 895–898, Amsterdam, 1998. IOS Press.

[111] Marina L. Kleptsyna, Alain Le Breton, and Michel Viot. Separation principle in the fractional Gaussian linear-quadratic regulator problem with partial observation. *ESAIM & PS*, 12:94–126, 2008.

[112] Urs Köster and Aapo Hyvärinen. A two-layer ICA-like model estimated by score matching. In Joaquim Marques de Sá, Luís A. Alexandre, Włodzisław Duch, and Danilo Mandic, editors, *International Conference on Artificial Neural Networks (ICANN 2007)*, volume 4669 of *LNCS*, pages 798–807. Springer, Heidelberg, 2007.

[113] Tuvia Kotzer, Nir Cohen, and Joseph Shamir. Generalized projection algorithms with applications to optics and signal restoration. *Optics Communications*, 156(1):77–91, 1998.

[114] P. Krishnaiah and Jugan Lin. Complex elliptically symmetric distributions. *Communications in Statistics*, 15(12):3693–3718, 1986.

[115] Erik G. Learned-Miller and John W. Fisher III. ICA using spacings estimates of entropy. *Journal of Machine Learning Research*, 4:1271–1295, 2003.

[116] Intae Lee and Te-Won Lee. Nonparametric independent component analysis for circular complex variables. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, volume 2, pages 665–668, Honolulu, Hawaii, USA, 15-20 April 2007.

[117] David N. Levin. Using state space differential geometry for nonlinear blind source separation. Technical report, University of Chicago, 2006. http://arxiv.org/abs/cs/0612096.

[118] Stan Z. Li, XiaoGuang Lv, and HongJiang Zhang. View-based clustering of object appearances based on independent subspace analysis. In *International Conference on Computer Vision (ICCV'01)*, volume 2, pages 295–300, Vancouver, BC, Canada, 2001.

[119] Stan Z. Li, XiaoGuang Lv, and HongJiang Zhang. View-subspace analysis of multi-view face patterns. In *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01)*, pages 125–132, Vancouver, BC, Canada, 2001.

[120] Andrew E. B. Lim, John B. Moore, and Leonid Faybusovich. Separation theorem for linearly constrained LQG optimal control. *Systems & Control Letters*, 28(4):227–235, 1996.

[121] Andrew E. B. Lim, John B. Moore, and Leonid Faybusovich. Separation theorem for linearly constrained LQG optimal control - continuous time case. In *Proceedings of the 35th IEEE Conference on Decision and Control*, volume 4, pages 4152–4157, 1996.

[122] Ross A. Lippert. *Nonlinear Eigenvalue Problems.* PhD thesis, Massachusetts Institute of Technology, 1998.

[123] David G. Luenberger. An introduction to observers. *IEEE Transactions on Automatic Control AC-16*, pages 596–602, 1971.

[124] Christophe De Luigi and Eric Moreau. Optimal joint diagonalization of complex symmetric third-order tensors. Application to separation of non circular signals. In Mike E. Davies, Christopher J. James, Samer A. Abdallah, and Mark D. Plumbley, editors, *Independent Component Analysis and Signal Separation (ICA 2007)*, volume 4666 of *LNCS*, pages 25–32. Springer, Heidelberg, 2007.

[125] Adam MacDonald and Stephen Cain. Derivation and application of an anisoplanatic optical transfer function for blind deconvolution of laser radar imagery. *Unconventional Imaging*, 5896:9–20, 2005.

[126] Manfredi Maggiore and Kevin Passino. A separation principle for a class of non-UCO systems. *IEEE Transactions on Automatic Control*, 48(7):1122–1133, July 2003.

[127] Scott Makeig, Anthony J. Bell, Tzyy-Ping Jung, and Terrence J. Sejnowski. Independent component analysis of electroencephalographic data. In *Neural Information Processing Systems (NIPS '96)*, volume 8, pages 145–151, 1996.

[128] Anke Meyer-Bäse, Peter Gruber, Fabian J. Theis, and Simon Foo. Blind source separation based on self-organizing neural network. *Engineering Applications of Artificial Intelligence*, 19:305–311, 2006.

[129] Terence C. Mills. *Time Series Techniques for Economists*. Cambridge University Press, Cambridge, 1990.

[130] Nikolaos Mitianoudis and Michael E. Davies. Audio source separation of convolutive mixtures. *IEEE Transactions on Speech and Audio Processing*, 11(5):489–497, September 2003.

[131] Eric Moreau. An any order generalization of JADE for complex source signals. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP 2001)*, volume 5, pages 2805–2808, May 2001.

[132] Gou Nakura and Akira Ichikawa. Stabilization of a nonlinear jump system. *Systems & Control Letters*, 47:79–85, 2002.

[133] Arnold Neumaier and Tapio Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, 27(1):27–57, 2001.

[134] Yasunori Nishimori, Shotaro Akaho, Samer A. Abdallah, and Mark D. Plumbley. Flag manifolds for subspace ICA problems. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, volume 4, pages 1417–1420, Honolulu, Hawaii, USA, 15-20 April 2007.

[135] Yasunori Nishimori, Shotaro Akaho, and Mark D. Plumbley. Riemannian optimization method on the flag manifold for independent subspace analysis. In Justinian Rosca, Deniz Erdogmus, José C. Príncipe, and Simon Haykin, editors, *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *LNCS*, pages 295–302. Springer, 5-8 March 2006.

[136] Yasunori Nishimori, Shotaro Akaho, and Mark D. Plumbley. Riemannian optimization method on the generalized flag manifold for complex and subspace ICA. In *MaxEnt 2006: International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, CNRS*, Paris, France, 8-13 July 2006.

[137] Guido Nolte, Frank C. Meinecke, Andreas Ziehe, and Klaus-Robert Müller. Identifying interactions in mixed and noisy complex systems. *Physical Review E*, 73(051913), 2006.

[138] Michael Novey and Tülay Adali. Adaptable nonlinearity for complex maximization of nongaussianity and a fixed-point algorithm. In *IEEE Workshop on Machine Learning for Signal Processing (MLSP 2006)*, pages 79–84, Maynooth, Ireland, September 2006.

[139] Michael Novey and Tülay Adali. Complex fixed-point ICA algorithm for separation of QAM sources using gaussian mixture model. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, volume 2, pages 445–448, Honolulu, Hawaii, USA, 15-20 April 2007.

[140] Simon Osindero, Max Welling, and Geoffrey E. Hinton. Topographic product models applied to natural scene statistics. *Neural Computation*, 18(2):381 – 414, 2006.

[141] Barak A. Pearlmutter and Lucas C. Parra. Maximum likelihood blind source separation: A context-sensitive generalization of ICA. In M.C. Mozer, M.I. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems (NIPS '97)*, volume 9, pages 613–619. MIT press, 1997.

[142] Michael S. Pedersen, Jan Larsen, Ulrik Kjems, and Lucas C. Parra. A survey of convolutive blind source separation methods. In *Springer Handbook of Speech Processing*. Springer Press, November 2007.

[143] V.V. Petrov. Central limit theorem for m-dependent variables. In *Proceedings of the All-Union Conference on Probability Theory and Mathematical Statistics*, pages 38–44, 1958.

[144] Mark D. Plumbley. Lie group methods for optimization with orthogonality constraints. In Carlos García Puntonet and Alberto Prieto, editors, *Independent Component Analysis and Blind Signal Separation (ICA 2004)*, volume 3195 of *LNCS*, pages 1245–1252. Springer, 2004.

[145] Barnabás Póczos and András Lőrincz. Independent subspace analysis using geodesic spanning trees. In *International Conference on Machine Learning (ICML 2005)*, volume 119, pages 673–680, New York, NY, USA, 2005. ACM Press.

[146] Barnabás Póczos and András Lőrincz. Independent subspace analysis using k-nearest neighborhood distances. *Artificial Neural Networks: Formal Models and their Applications (ICANN 2005)*, 3697:163–168, 2005.

[147] Barnabás Póczos and András Lőrincz. Non-combinatorial estimation of independent autoregressive sources. *Neurocomputing Letters*, 69:2416–2419, 2006.

[148] Barnabás Póczos, Bálint Takács, and András Lőrincz. Independent subspace analysis on innovations. In *European Conference on Machine Learning (ECML 2005)*, volume 3720 of *LNAI*, pages 698–706. Springer-Verlag, 2005.

[149] Nicolas Quinquis, Isao Yamada, and Kohichi Sakaniwa. Efficient dual Cayley parametrization technique for ICA with orthogonality constraints. In *ICA Research Network International Workshop (ICARN 2006)*, pages 123–126, Liverpool, U.K., 18-19 September 2006.

[150] Ravikiran Rajagopal and Lee C. Potter. Multivariate MIMO FIR inverses. *IEEE Transactions on Image Processing*, 12:458 – 465, 2003.

[151] Raymond Rishel. A strong separation principle for stochastic control systems driven by a hidden markov model. *SIAM Journal on Control and Optimization*, 32(4), 1994.

[152] Michael J. Roan, Mark R. Gramann, Josh G. Erling, and Leon H. Sibul. Blind deconvolution applied to acoustical systems identification with supporting experimental results. *The Journal of the Acoustical Society of America*, 114(4):1988–1996, October 2003.

[153] Reuven Y. Rubinstein and Dirk P. Kroese. *The Cross-Entropy Method*. Springer, 2004.

[154] Tapio Schneider and Arnold Neumaier. Algorithm 808: ARfit - a matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, 27(1):58–65, 2001.

[155] Alok Sharma and Kuldip K. Paliwal. Subspace independent component analysis using vector kurtosis. *Pattern Recognition*, 39:2227–2232, 2006.

[156] Hao Shen and Knut Hüper. Generalised FastICA for independent subspace analysis. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, volume 4, pages 1409–1412, Honolulu, Hawaii, USA, 15-20 April 2007.

[157] Hao Shen, Knut Hüper, and Martin Kleinsteuber. Block-Jacobi-type methods for log-likelihood based linear independent subspace analysis. In *IEEE International Workshop on Machine Learning (IEEE MLSP07)*, pages 133–138, Thessaloniki, Greece, 27-29 August 2007.

[158] Herbert A. Simon. Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica*, 24:74–81, 1956.

[159] Paris Smaragdis. Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, 22:21–34, 1998.

[160] Jordi Solé-Casals, Christian Jutten, and Dinh Tuan Pham. Fast approximation of nonlinearities for improving inversion algorithms of PNL mixtures and Wiener systems. *Signal Processing*, 85:1780–1786, 2005.

[161] Harald Stögbauer, Alexander Kraskov, Sergey A. Astakhov, and Peter Grassberger. Least dependent component analysis based on mutual information. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 70(066123), December 2004.

[162] V. Sundarapandian. A separation theorem for robust pole placement of discrete-time linear control systems with full-order observers. *Mathematical and Computer Modelling*, 43:42–48, 2006.

[163] Ananthram Swami, Georgios Giannakis, and Sanyogita Shamsunder. Multichannel arma processes. *IEEE Transactions on Signal Processing*, 42(4):898–913, April 1994.

[164] Seiji Takano. The inequalities of Fisher information and entropy power for dependent variables. In *Proceedings of the 7th Japan-Russia Symposium on Probability Theory and Mathematical Statistics*, July 1995.

[165] Anisse Taleb and Christian Jutten. Source separation in post-nonlinear mixtures. *IEEE Transactions on Signal Processing*, 10(47):2807–2820, October 1999.

[166] Inder Jeet Taneja. *Generalized Information Measures and Their Applications*. on-line book: www.mtm.ufsc.br/~taneja/book/book.html, 2001.

[167] Henri Theil. A note on certainty equivalence in dynamic planning. *Econometrica*, 25:346–349, 1957.

[168] Fabian J. Theis. A new concept for separability problems in source separation. *Neural Computation*, 16:1827–1850, 2004.

[169] Fabian J. Theis. Uniqueness of complex and multidimensional independent component analysis. *Signal Processing*, 84(5):951–956, 2004.

[170] Fabian J. Theis. Blind signal separation into groups of dependent signals using joint block diagonalization. In *International Society for Computer Aided Surgery (ISCAS 2005)*, volume 6, pages 5878–5881, 2005.

[171] Fabian J. Theis. Multidimensional independent component analysis using characteristic functions. In *European Signal Processing Conference (EUSIPCO 2005)*, 2005.

[172] Fabian J. Theis. Towards a general independent subspace analysis. In *Neural Information Processing Systems (NIPS 2006)*, volume 19, 2006.

[173] Fabian J. Theis and Motoaki Kawanabe. Uniqueness of non-gaussian subspace analysis. In Justinian Rosca, Deniz Erdogmus, José C. Príncipe, and Simon Haykin, editors, *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *Lecture Notes in Computer Science*, pages 917–925. Springer, 5-8 March 2006.

[174] Fabian J. Theis and Motoaki Kawanabe. Colored subspace analysis. In Mike E. Davies, Christopher J. James, Samer A. Abdallah, and Mark D. Plumbley, editors, *Independent Component Analysis and Signal Separation (ICA 2007)*, volume 4666 of *LNCS*, pages 121–128. Springer, Heidelberg, 2007.

[175] James V. Uspensky. Asymptotic formulae for numerical functions which occur in the theory of partitions. *Bulletin of the Russian Academy of Sciences*, 14(6):199–218, 1920.

[176] Ricardo Vigário, Veikko Jousmäki, Matti Hämäläinen, Riitta Hari, and Erkki Oja. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Neural Information Processing Systems (NIPS '97)*, volume 10, pages 229–235, Cambridge, MA, 1997.

[177] Roland Vollgraf and Klaus Obermayer. Multi-dimensional ICA to separate correlated sources. In *Neural Information Processing Systems (NIPS 2001)*, volume 14, pages 993–1000, Cambridge, MA, 2001. MIT Press.

[178] Cabir Vural and William A. Sethares. Blind image deconvolution via dispersion minimization. *Digital Signal Processing*, 16:137–148, 2006.

[179] W.M. Wonham. On the separation theorem of stochastic control. *SIAM Journal on Control*, 6(2):312–326, 1968.

[180] Jian-Wu Xu, Deniz Erdogmus, Yadunandana N. Rao, and José Carlos Príncipe. Minimax mutual information approach for ICA of complex-valued linear mixtures. In *Independent Component Analysis and Blind Signal Separation (ICA 2004)*, volume 3195 of *LNCS*, pages 311–318, 22-24 September 2004.

[181] Arie Yeredor. Blind separation of gaussian sources via second-order statistics with asymptotically optimal weighting. *IEEE Signal Processing Letters*, 7(7), 2000.

[182] Stella X. Yu and Jianbo Shi. Multiclass spectral clustering. In *International Conference on Computer Vision (ICCV 2003)*, volume 1, pages 313–319. IEEE Computer Society, October 2003.

[183] Andreas Ziehe, Motoaki Kawanabe, Stefan Harmeling, and Klaus-Robert Müller. Blind separation of post-nonlinear mixtures using linearizing transformations and temporal decorrelation. *Journal of Machine Learning Research*, 4(7-8):1319–1338, 2004.